AD_____

Award Number: DAMD17-00-1-0372

TITLE: Selective Inhibitors of 17β-Hydroxysteroid Dehydrogenase

PRINCIPAL INVESTIGATOR: David L. Vander Jagt, Ph.D.

CONTRACTING ORGANIZATION: University of New Mexico
Albuquerque, New Mexico 87131

REPORT DATE: July 2003

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	July 2003	Final (15 Jun 2000 - 14 Jun 2003)

4. TITLE AND SUBTITLE
Selective Inhibitors of 17β-Hydroxysteroid Dehydrogenase

5. FUNDING NUMBERS
DAMD17-00-1-0372

6. AUTHOR(S)
David L. Vander Jagt, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of New Mexico
Albuquerque, New Mexico 87131

E-Mail: dlvanderjagt@salud.unm.edu

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 Words)

17β-Hydroxysteroid dehydrogenase type I (HSD-I) was purified from human placenta and was used to screen a set of gossypol derivatives and to identify two lead compounds (gossylic lactone and gossylic iminolactone) for structure based drug design. Gossylic lactone and gossylic iminolactone were synthesized. The schemes were then modified to prepare analogs in the hemi-gossylic family. Synthetic work is in progress to prepare a number of compounds, based upon computational chemistry and molecular modeling. Key intermediates in the synthetic schemes have been prepared. Modeling results are in agreement with the hypothesis of this study that the Rossmann fold is the target of these inhibitors. This is also consistent with the observation that the lead compounds are competitive inhibitors of the binding of cofactor to HSD-I. Modeling studies have been used to design Pan-Active Site inhibitors that are predicted to span both the cofactor and substrate sites. Development of a new molecular docking algorithm was completed. A publication from this work is attached along with a submitted publication

14. SUBJECT TERMS
Polypyrimidine Tract Binding Protein (PTB), Fibroblast Growth Factor Receptor 2 (FGFR2), Alternative Splicing, Fox-1

15. NUMBER OF PAGES
45

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

Table of Contents

INTRODUCTION

Human Type I 17β-HSD, also known as 17β-estradiol dehydrogenase, catalyzes the reduction of the weak estrogen, estrone, to the strong estrogen, 17β-estradiol, which is the biologically active estrogen involved in the development of human breast cancer. Type I 17β-HSD is therefore a very attractive target for drug development.

Objectives: Recently, we developed a new class of dehydrogenase inhibitors that are targeted at the NAD(P)/NAD(P)H binding sites (Rossmann fold) of dehydrogenses. Surprisingly, these inhibitors exhibit selectivity for different dehydrogenases. The goal of this project is to develop selective inhibitors of human Type I 17β-HSD as "lead compounds" for structure-based drug design. The crystal structure of human Type 1 17β–HSD is available to aid in structure-based drug design. The concept that the Rossmann fold may represent a useful drug target is a new concept in drug design.

Specific Aims: Specific Aim 1: To develop versatile synthetic schemes for the preparation of a wide range of substituted hydroxynaphthoic acids as potential dehydrogenase inhibitors, utilizing the principles of convergent synthesis and combinatorial chemistry to prepare libraries of compounds; Specific Aim 2: To utilize molecular modeling, kinetics, fluorescence quenching studies and crystallography for design of HSD inhibitors using classical drug design/optimization methods.

Approved Work Plan

Task 1
Purification of 17β–hydroxysteroid dehydrogenase from human placenta (months 1-3)

Task 2
Development of synthetic schemes for preparation of mono- and dihydroxynaphthoic acids as potential inhibitors of 17β-HSD-1 (months 1-36, an ongoing activity)

Task 3
Development of combinatorial libraries (months 1-36, an ongoing activity)

Task 4
Development of Pan-Active-Site inhibitors, directed by molecular modeling and kinetic results (months 1-36, an ongoing activity)

Task 5
Development of enzyme assays, kinetic procedures, fluorescence quenching procedures (months 1-3)

Task 6
Development of molecular modeling procedures (months 1-12, and used thereafter on regular basis)

Task 7
Cell culture studies of human breast cancer cells (months 12-36)

BODY

I. 17β-Hydroxysteroid Dehydrogenase: Purification from Placenta

Methods (1): The purification of placental 17β-hydroxysteroid dehydrogenase type I (HSD) was carried out utilizing a rapid and efficient purification scheme. Routinely approximately 100 gms of snap-frozen cubed placenta was homogenized in a blender in 250 ml cold homogenizing buffer for 5 minutes on ice. A cocktail of protease inhibitors (antipain, bestatin, chymostatin, pepstatin A, and leupeptin) was added to a final concentration of 0.5 μg/ml. Next, the sample was sonicated (on ice)

three times 60 seconds at 50 watts using a Sonifier Cell Disruptor, Model W185D (Heat Systems-Ultrasonics, Inc., Plainview, NY) fitted with a large sonicating probe. The solution was then centrifuged at 700 x g for 15 minutes and then 6,000 x g for 15 minutes This supernatant was further centrifuged at 100,000 x g for 30 minutes.

Normally 24 hours before purification of HSD, the 250 ml Blue Sepharose CL-6B column is cleaned and regenerated by washing with 500 ml of; Millipore filtered water, 4.5 M urea, 0.1 M tris-HCl buffer pH 8.5 containing 0.5 M NaCl, 0.1 M sodium acetate buffer pH 4.5 containing 0.5 M NaCl. Finally, the column is brought to pH 7.4 with PBS.

The Blue Sepharose column was equilibrated with 40 mM tris-HCl buffer pH 7.5 . Immediately after the 100,000 x g centrifugation was complete. The supernatant was loaded onto the Blue Sepharose CL-6B column. The column was then washed thoroughly with 300 ml of buffer. Finally the HSD was eluted from the column with buffer containing 5 mM NAD. The eluent was concentrated to 7.5 ml by pressure filtration in an Amicon apparatus fitted with a YM-10 membrane. The concentrated HSD was desalted on a PD-10 column which had been equilibrated with desalting buffer.

The desalted HSD sample was immediately loaded onto the chromatofocusing column (8 mm x 155 mm) of PBE 94 resin equilibrated with 25 mM imidazole buffer pH 7 (PBE-94 buffer). The chromatofocusing column was developed with 54 ml of polybuffer 74 (1:8 dilution with Millipore filtered water) at pH 4 (polybuffer).

The pHs were taken on a Radiometer pH meter, type PHM26. The HSD activity was determined utilizing a Perkin/Elmer Lambda 6 spectrophotometer. Activity was assayed in 1 ml total volume of 0.1 M sodium bicarbonate buffer pH 9.2, 25 mM estradiol, and 0.5 mM NAD at 340 nm. ($\varepsilon = 6.2^{-1}$ cm^{-1}).

The HSD sample was almost homogeneous (>95%) at this point as shown by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), 20% with a 5% stacking gel which was developed by silver stain The SDS-PAGE showed only slight contamination with the observance of one additional band of a higher molecular weight protein.

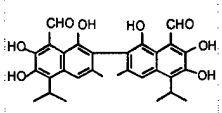
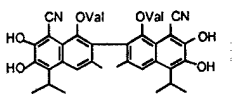
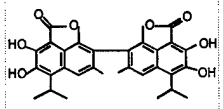
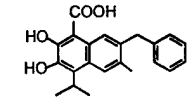
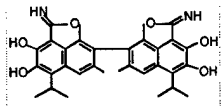
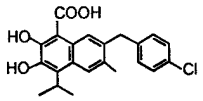
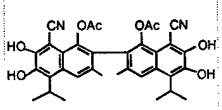
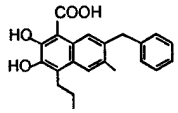
If continued purification is desired it can be accomplished utilizing hydroxylapatite.

Continued purification of β-HSD was carried out utilizing hydroxylapatite (HPHT). A few grams of HPHT (Bio-Rad) was added to approximately 100 ml of Millipore filtered water and gently swirled to perturb the "fines" up into the water. The major HPHT crystals were allowed to settle and the "fines" were poured off. This procedure was repeated several times until the HPHT appeared to settle out quickly and no "fines" were observed in the water. A column (8 mm x 80 mm) was meticulously poured continuously. The column was equilibrated with PBE-94 buffer . The chromatofocused HSD was again concentrated on Amicon YM-10 to approximately 7.5 mls and desalted through a PD-10 column pre-equilibrated with buffer as above). This desalted solution was loaded onto the HPHT column. A linear gradient from 10 mM sodium phosphate buffer pH 6.8 containing 0.1 M NaCl to 350 mM sodium phosphate buffer pH 6.8 containing 0.1 M NaCl was run in 80 ml total volume. Fractions were collected at 2 ml/tube. The peak tubes were pooled. The enzyme was stored at -70°C in storage buffer.

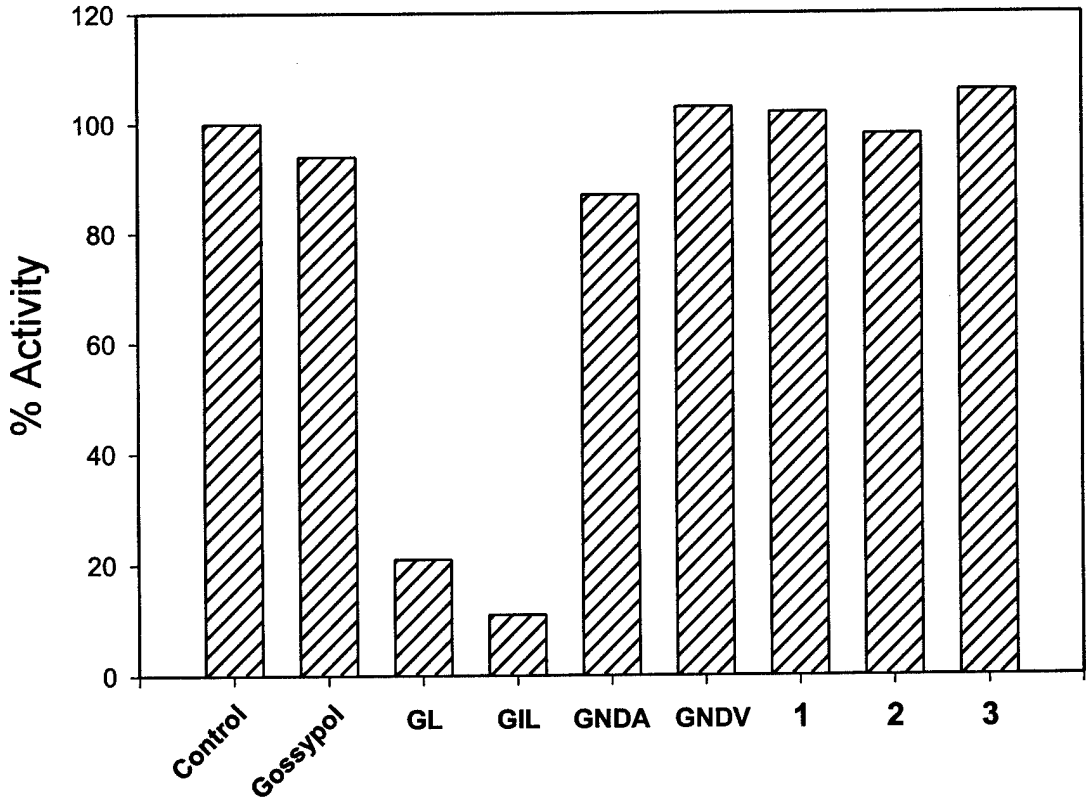
II. Identification of Inhibitors of HSD as Lead Compounds for Structure-based Drug Design

A series of compounds related to the natural product gossypol was screened against HSD-1 (figure 1). Two compounds, gossylic lactone (GL) and gossylic iminolactone (GIL) were selected as lead compounds. Both GL and GIL are competitive inhibitors of the binding of cofactor as shown in figure 2 Ki values are 2.2 and 4.3 micromolar for GL and GIL, respectively.

Gossypol and Related Compounds (GRCs)

Gossypol	GS		**Gossylic Nitrile 1,1-Divalerate**	GND V	
Gossylic Lactone	GL		**7-benzyl-8-deoxyhemigossylic acid**	**1**	
Gossylic Iminolactone	GIL		**7-p-chloro-8-deoxyhemigossylic acid**	**2**	
Gossylic Nitrile 1,1-Diacetate	GND A		**2,3-dihydroxy-6-methyl-4-propylnapthoic acid**	**3**	

17β-Hydroxysteroid Dehydrogenase from Human Placenta



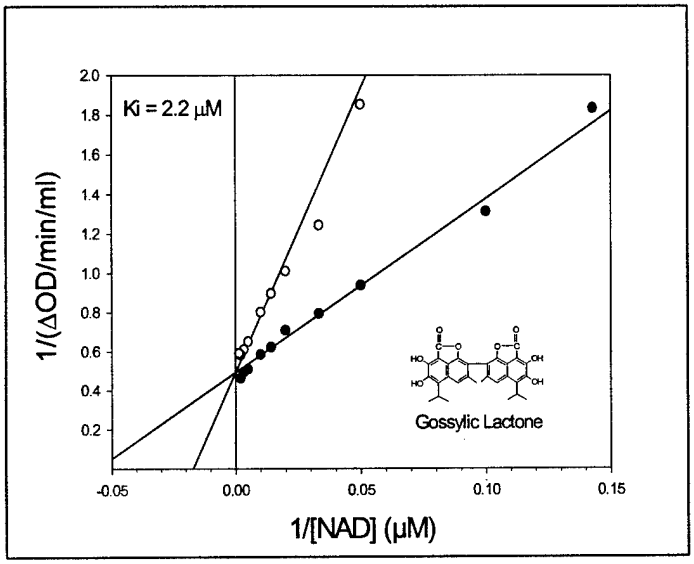
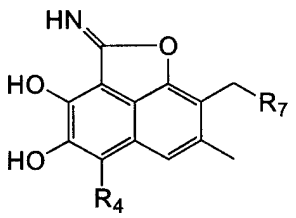


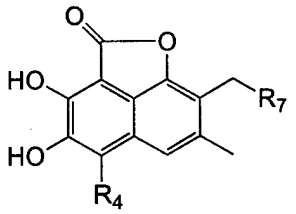
Figure 2

III. Synthesis

The synthetic work has been directed toward an efficient synthesis of monomeric compounds analogous to gossylic iminolactone and gossylic lactone with the following general structures:



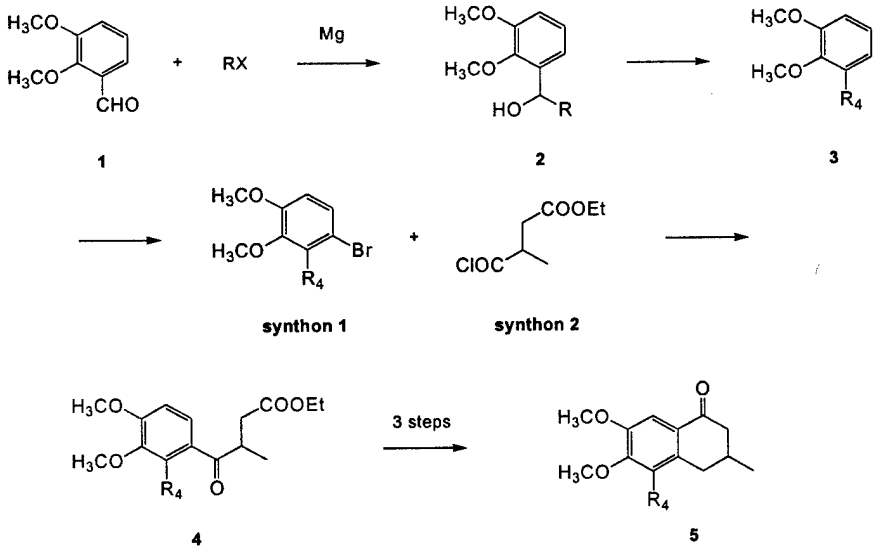
Iminolactone



Lactone

These compounds will have various groups at the positions labeled R_4 and R_7. The R_4 groups are introduced early in the synthesis as indicated in Scheme I.

Scheme I

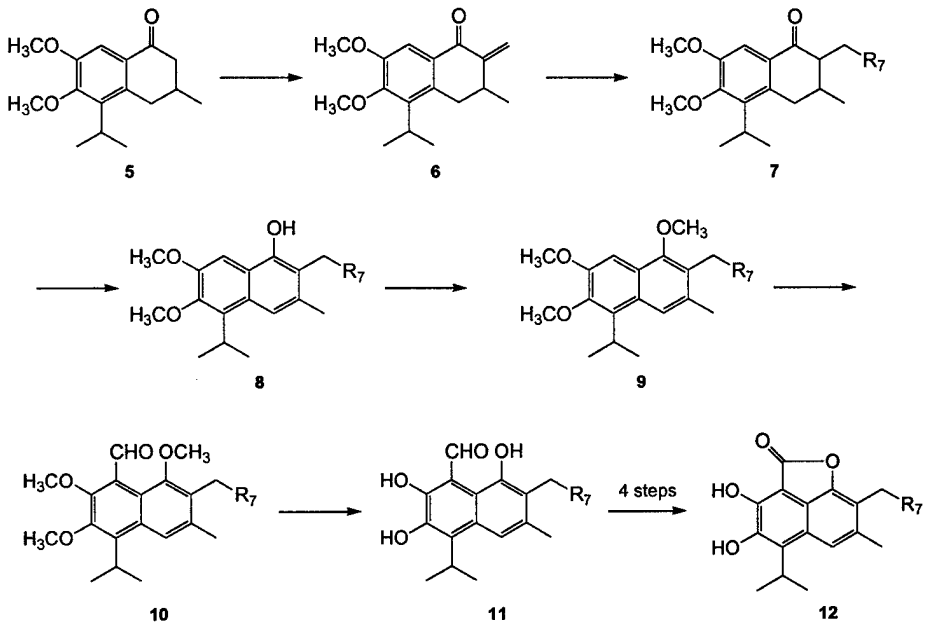


When the two halves of the molecule (synthon **1** and synthon **2**) are brought together by a Grignard reaction, the R_4 group is already incorporated in synthon **1**. Synthon **1** is made by the Grignard reaction of 2,3-dimethoxybenzaldehyde (**1**) with an alkyl halide. The resulting alcohol (**2**) is hydrogenolyzed to provide a dimethoxybenzene with R_4 incorporated (**3**). Bromination of **3** provides synthon **1**. The R_4 groups which have been incorporated thus far include methyl, ethyl, propyl, isopropyl (by a different method) and butyl. Presently the synthesis of compounds with cyclopentylbutyl and methoxyethyl R_4 groups is in progress.

Efforts to improve the overall synthesis by incorporating aldehyde equivalent groups into synthon **1** have not proven fruitful nor have efforts to incorporate R_7 into synthon **2**. However, the synthesis was improved by using an acid chloride rather than an aldehyde as the reactive functionality of synthon **2**.

The synthetic procedures being developed to place various groups at position R_7 are shown in Scheme II (where the R_4 group is shown as isopropyl).

Scheme II



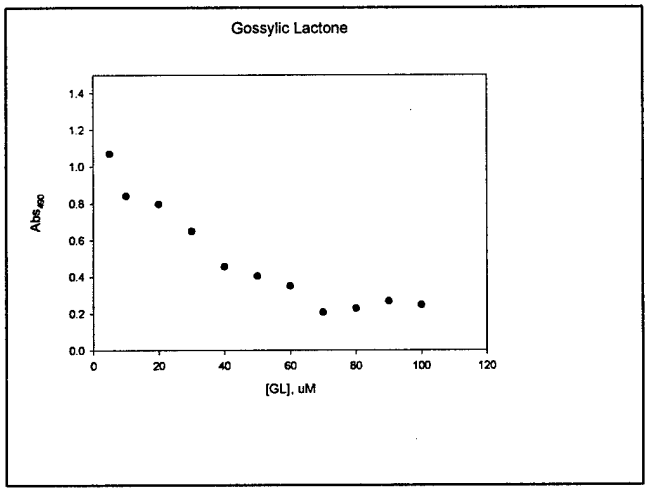
Numerous attempts to alpha alkylate tetralone **5** with a saturated alkyl group proved to be inefficient, always resulting in a large amount of dialkylated product. Therefore, the methylenation of **5,** shown in Scheme II, was developed. It is anticipated that various R_7 groups can be introduced by Michael addition to unsaturated ketone **6** to prepare a variety compounds. Thus far, compound **6** has been reduced to the saturated tetralone **7** (R_7 = methyl), aromatized to form the corresponding phenol (**8**) and methylated to form the trimethoxynapthalene **9**. Compound **9** will be formylated with t-butyllithium and dimethylformamide. The rest of the synthesis will be accomplished using procedures already worked out in this laboratory.

IV. Molecular Modeling (see appendix); Abstract from Chemico Biological Interact 143-144, 481-491 (2003): 17-β-Hydroxysteroid dehydrogenase type 1 (17βHSD1), also called estradiol dehydrogenase, catalyzes the NADPH-dependent reduction of the weak estrogen, estrone, into the more potent estrogen, 17-β-estradiol. 17βHSD1 is an attractive drug target in hormone sensitive breast cancer. Past efforts to develop selective inhibitors of 17βHSD1 have focused on design of substrate analogs. It is challenging to develop steroid analogs that are devoid of any undesired biological activity. 17βHSD1 is a member of the short chain dehydrogenase/reductase (SDR) superfamily that includes many hydroxysteroid dehydrogenases. Members of the SDR family bind NAD(P)(H) in a motif that is a modified Rossmann fold. We demonstrated previously that the Rossmann folds of classical dehydrogenases can be selectively inhibited by derivatives and analogs of the

natural product gossypol. In the present study, we have addressed the question whether the modified Rossmann fold in 17βHSD1 is a target for identification of lead compounds for structure-based drug design. 17βHSD1 was purified from human placenta. 17βHSD1 is inhibited by derivatives of gossypol with dissociation constants as low as 2 micromolar. Inhibition is competitive with the binding of NADPH. Molecular modeling studies (AutoDock 3.0) using the published coordinates of human 17βHSD1 suggest that these inhibitors occupy the modified Rossmann fold at the nicotinamide end of the NADPH-binding site, extending towards the substrate site. A computational approach was used to design potential new inhibitors of 17βHSD1. The results suggest not only that derivatives of gossypol represent attractive lead compounds for structure-based drug design but also suggest that appropriate incorporation of a substrate analog into the design of these Rossmann fold inhibitors may provide Pan-Active Site inhibitors that span the cofactor and substrate site, potentially offering specificity and increased potency.

V. New Algorithms (see appendix); Abstract from J Comp Chem, submitted: A new approach for defining the Cartesian spatial boundaries of binding pockets is presented. The method involves calculation of a *macromolecule encapsulating surface* (MES) that separates binding pocket volume from outside space. The surface provides means for identification of binding sites and calculation of their volume. Additionally, the MES can be used to limit the search space for ligand docking and *de novo* design algorithms via identification of accessible atoms within the binding pocket or limitation of translation ranges to binding pocket space. The approach has been shown to be efficacious based on testing with 50 enzyme-ligand complexes for which the binding pockets are known. Additionally, we have modified the flexible docking program AutoDock 3.0 to incorporate MES boundaries using an energetic term. The results show increased efficiency of the genetic algorithm for ligand docking characterized by a larger percentage of successful runs and a decrease in required run times. MES incorporation also facilitates search of an entire enzyme for ligand docking, without the requirement of a predetermined binding pocket location

VI. Cellular Studies



In preliminary cellular studies, the activity of gossylic lactone against the growth of human breast cancer cells (TSE) was determined. Gossylic lactone, which shows low toxicity against non-cancer cells, exhibited an IC_{50} of 25 micromolar.

KEY RESEARCH ACCOMPLISHMENTS

1. Identification of two lead compounds as inhibitors of HSD
2. Continued development of synthetic strategies to prepare second generation inhibitors of HSD
3. Molecular modeling studies to design Pan-Active Site inhibitors of HSD
4. Design of a new molecular modeling approach to drug design
5. Published paper and manuscript submitted describing the initial phase of this project
6. The activity of GL against breast cancer cell line TSE was demonstrated

REPORTABLE OUTCOMES

Compounds related to the natural product gossypol have been developed as lead compounds for the inhibition of human HSD. These compounds are competitive inhibitors of the binding of cofactor to the Rossmann fold. Molecular modeling studies are being used to prepare second generation inhibitors as potential drugs for treatment of breast cancer. A new modeling program has been incorporated into the design of new inhibitors.

W.M. Brown, R.E. Royer, L.M.Deck,L.A. Hunsaker and D.L. Vander Jagt, The Cofactor Site of Human 17-beta Hydroxysteroid Dehydrogenase Type I as a Drug Target. *FASEB J* 15, A1159 (2001)

William M. Brown, Louis E. Metzger, IV, Jeremy P. Barlow, Lucy A. Hunsaker, Lorraine M. Deck, Robert E. Royer, and David L. Vander Jagt, 17-β-Hydroxysteroid Dehydrogenase type 1: Computational Design of Active Site Inhibitors Targeted to the Rossmann Fold. *Chem Biol Interactions* 143-144, 481-491 (2003)

WM Brown and DL Vander Jagt, New Approach for Characterization of Binding Site Search Space. *J Comp Chem*, submitted (2003)

CONCLUSIONS

The results of this ongoing study are consistent with the hypothesis that the Rossmann fold of dehydrogenases can be exploited in the development of inhibitors of human HSD. In addition, the results suggest that Pan-Active Site inhibitors, in which a single molecule of the designed inhibitor will complex at both the cofactor and substrate binding sites, can be developed.

APPENDIX

William M. Brown, Louis E. Metzger, IV, Jeremy P. Barlow, Lucy A. Hunsaker, Lorraine M. Deck, Robert E. Royer, and David L. Vander Jagt, 17-β-Hydroxysteroid Dehydrogenase type 1: Computational Design of Active Site Inhibitors Targeted to the Rossmann Fold. *Chem Biol Interactions* 143-144, 481-491 (2003)

WM Brown and DL Vander Jagt, New Approach for Characterization of Binding Site Search Space. *J Comp Chem*, submitted (2003)

Addendum to final report for DAMD17-00-1-0372

PI: David L. Vander Jagt

Institution: University of New Mexico

Title: Selective Inhibitors of 17β-Hydroxysteroid Dehydrogenase

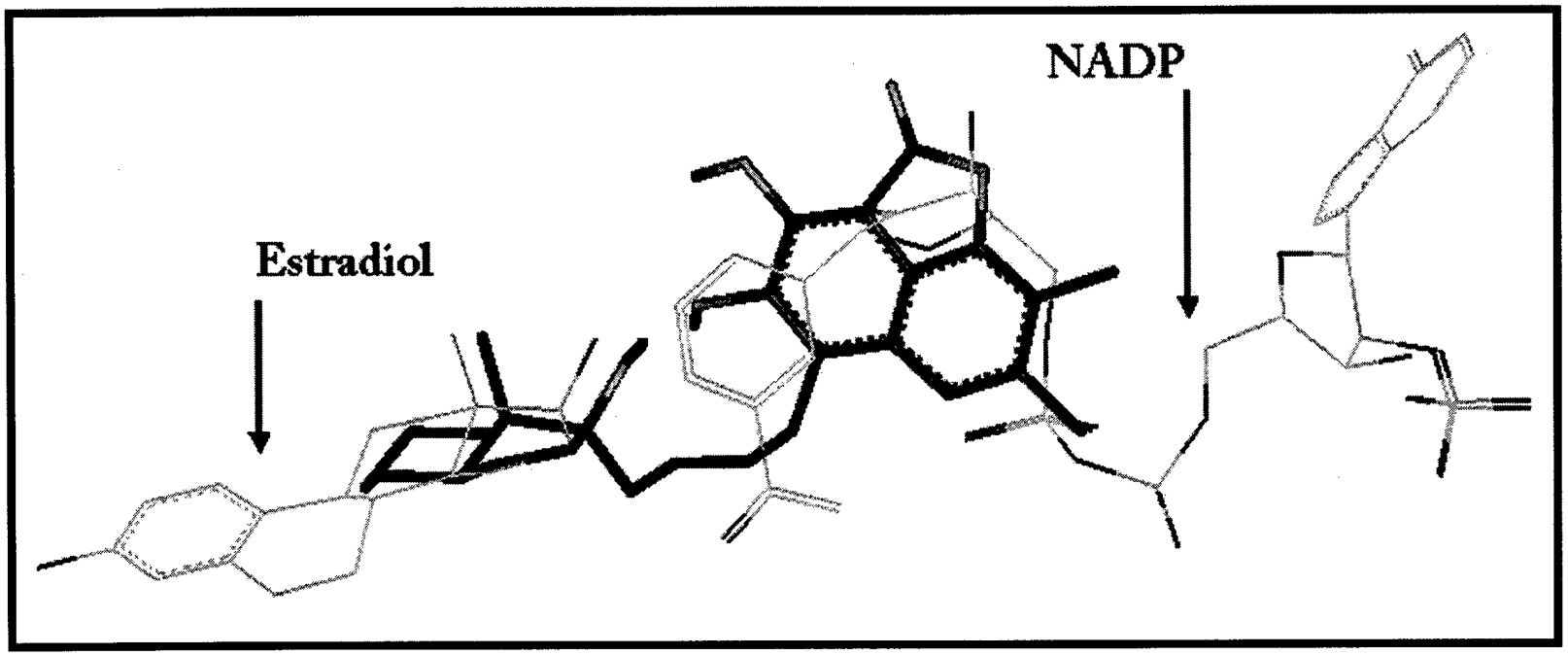
Summary: The final progress report was considered unacceptable because problems associated with completion of the objectives were not included. The attached material is being submitted in response to this criticism.

It should be noted that the synthesis of the target compound shown below (and included in the published paper titled "17-β–Hydroxysteroid dehydrogenase type 1: computational design of active site inhibitors targeted to the Rossmann fold", Chem Biol Interact **143-144,** 481-491 (2003)) initially involved synthetic chemistry (Scheme 1 in the attached materials) that utilized procedures reported in the literature, which were expected to be successful in preparing a hemi-lactone. This chemistry did not work. This required us to develop new synthetic chemistry, which is summarized in the attached material in Scheme II and Scheme III.

The target Pan-Active Site inhibitor shown below was based upon the initial observation that gossylic lactone was a good lead compound, and based upon our computational studies. We previously developed and reported on the development of versatile synthetic schemes to prepare dihydroxynaphthoic acids with different groups replacing the isopropyl group in the 4-position. This is the position where a substrate analog of estradiol will be introduced in the synthesis of the target molecule below. (This corresponds to the R_4 group shown in Scheme 1 of the original final report). This reported chemistry (J Med Chem **38**, 2427-2432 (1995); J Med Chem **41**, 3879-3887 (1998); Current Med Chem **7**, 479-498 (2000)) will allow us to complete the synthesis of the target compound once we solve the problem of introducing the hemi-lactone functionality.

The inclusion of a graph of inhibition of growth of breast cancer cell line TSE (an ER^+ line) was simply to provide preliminary data that showed activity with gossylic lactone. It is hypothesized that the target compound will be more active than gossylic lactone, based upon computational chemistry that predicts tighter binding of the target inhibitor. Once the target compound is synthesized, it will be tested against a battery of ER^+ and ER^- standard breast cancer lines.

PAN-ACTIVE SITE TARGET MOLECULE

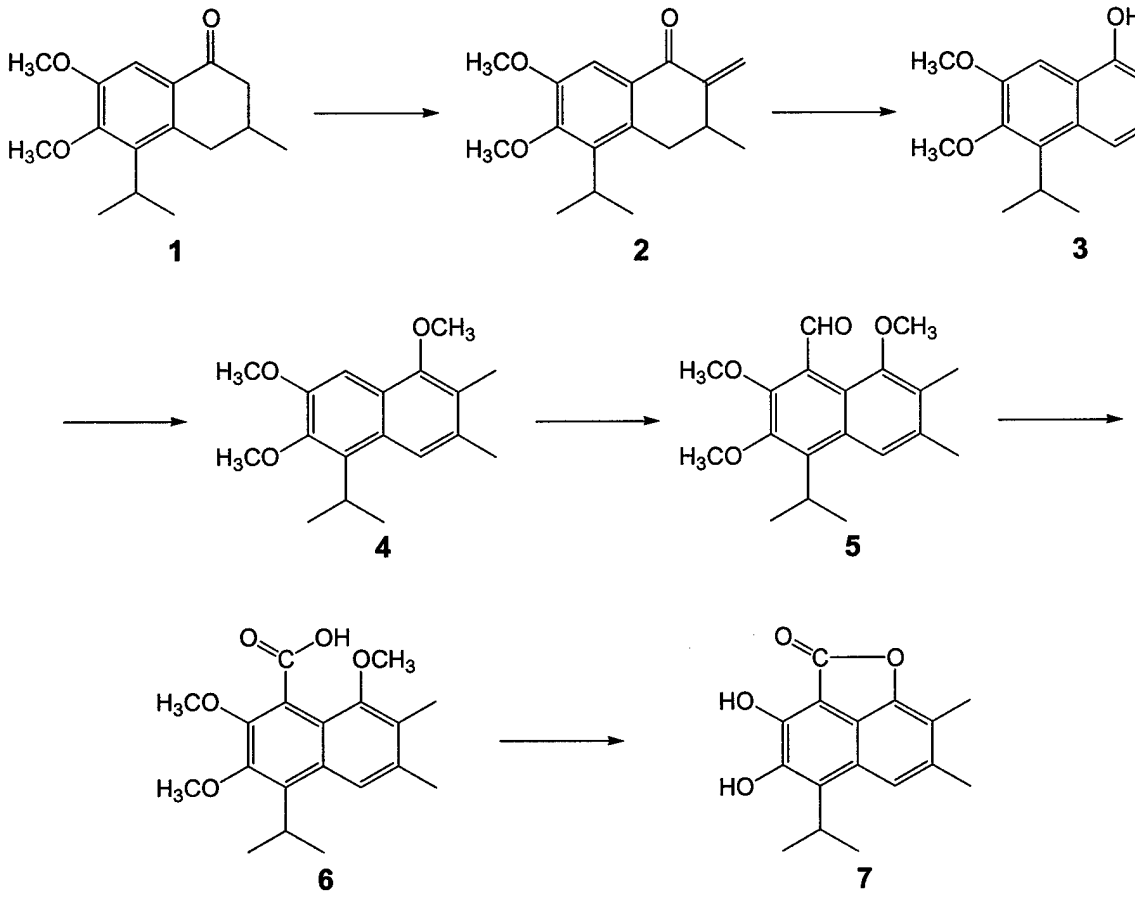


Pan-Active Site inhibitor target molecule that is predicted to occupy both the cofactor site of 17-β- HSD-type 1at the nicotinamide end of the Rossmann fold and extend into the estradiol binding site. The substrate analog of estradiol, which includes the C and D rings, is shown attached to the 4-position of the hemi-lactone analog of the lead inhibitor gossylic lactone.

Proposed Synthesis of Hemigossylic Lactones

The proposed synthesis of lactones related to hemigossypol is shown in Scheme I where the lactone form of 4-isopropyl-2,3,8-trihydroxy-6,7-dimethyl-1-naphthoic acid (**7**) is shown as an example. In other compounds of this type, the methyl group in the 7-position will be replaced with other groups and the isopropyl group in the 4-position will be replaced with an analog of estradiol.

Scheme I



Tetralone **1** is treated with paraformaldehyde and a catalyst to provide methylene substituted tetralone **2**.[1] Isomerization of **2** with a paladium catalyst in a high boiling solvent should provide phenol **3**. Compound **3** will be methylated to form the trimethoxynaphthalene **4**. Compound **4** will be formylated with t-butyl lithium and n-methylformanilide to provide aldehyde **5**.[2] The aldehyde group of **5** will be oxidized to the carboxylic acid group of **6**.[3] The methoxy methyl groups of **6** will be removed with boron tribromide to provide target compound **7**.

Problems with the Synthesis

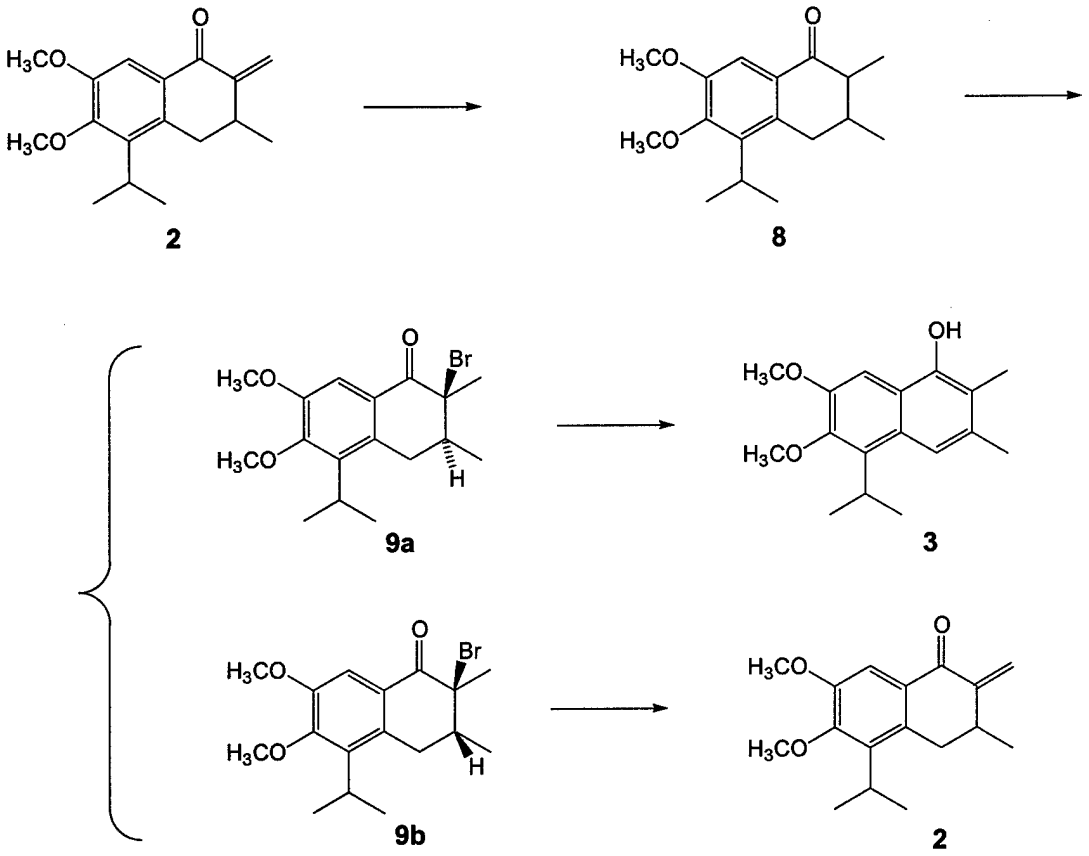
The intention was to isomerize **2** to **3** with a procedure reported in a patent for isomerizing, 2-methylene-1-oxo-1,2,3,4-tetrahydronaphthalene to 2-methyl-1-naphthol.[4] The procedure used

palladium on charcoal (pre-activated with hydrogen) in refluxing toluene. The procedure did not work for compound **2**.

Alternate synthesis of 3

Since the initial attempts to make **3** by isomerization of **2** failed, another route to **3** (shown in Scheme II) was tried.

Scheme II

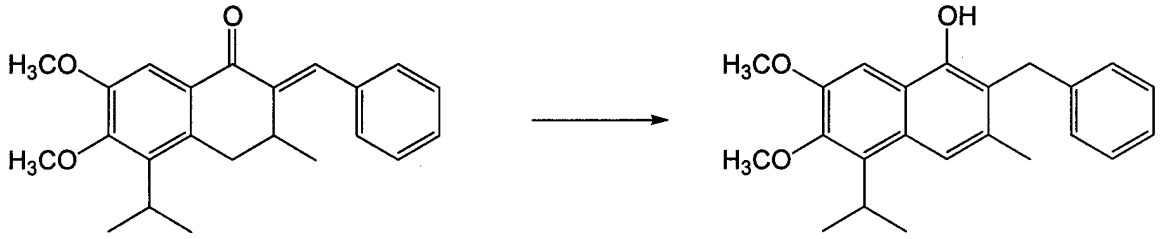


The exocyclic methylene group of compound **2** was reduced with palladium on charcoal in acetonitrile to form tetralone **8**. Compound **8** was brominated to form a mixture of the isomers **9a** and **9b** in approximately equal amounts. An attempt was made to dehydrobrominate the components of the mixture under various conditions. Isomer **9a**, in which the bromine and the hydrogen at the 3-position are in a trans configuration dehydrobrominated readily to form **3**. Isomer **9b** in which the bromine and the hydrogen at the 3-position are in a cis configuration was isolated from the product mixture and subjected to forcing dehydrobrominating conditions. It underwent an exocyclic elimination and reverted to starting material **2**. Since this approach requres two more steps than the direct isomerization of **2** and since the yield of the dehydrobromination step is less than 50%, the isomerization approach was reconsidered.

The literature procedure which seemed most promising was one reported in the Chinese literature which involves heating with palladium on charcoal at high temperature in ethylene glycol.[5] Thus far, this procedure has been tested on on benzylidene compound **10** (see Scheme

III) with the formation of benzyl substituted phenol **11**. It is anticipated that **2** can be converted to **3** under similar conditions. The reaction is presently under investigation.

Scheme III



10 **11**

References

1. J-L Gras (1980) Methylene ketones and aldehydes by simple, direct methylene transfer: 2-methylene-1-oxo-1,2,3,4-tetrahydronaphthalene. **Org Syn 60**, 88-91.

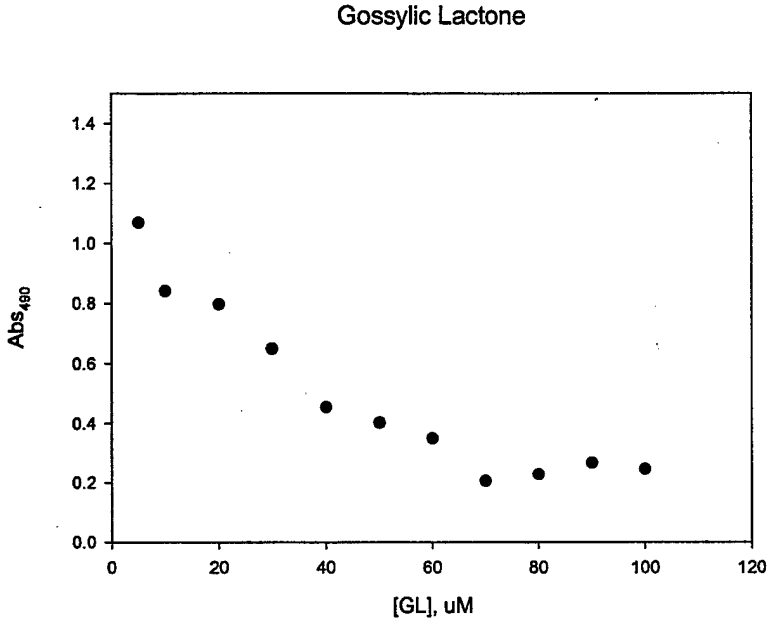
2. A Manmade, P Herlihy, J Quick, RP Duffley, M. Burgos, AP Hoffer (1983) Gossypol. Synthesis and in vitro spermicidal activity of isomeric hemigossypol derivatives. **Experientia 39**, 1276-1277.

3. E. Dalcanale, F. Montanari (1986) Selective oxidation of aldehydes to carboxylic acids with sodium chlorite – hydrogen peroxide. **J Org Chem 51**, 567-569.

4. Tamura and Tamai **EP 0 558 069 A1**

5. (1987) **Acta Chimica Sinica 45**, 506-509.

Figure Legend: Preliminary testing of the concept that an inhibitor of 17β-hydroxysteroid dehydrogenase type 1 will exhibit activity against an ER^+ breast cancer line (TSE). Gossylic lactone was selected as a lead inhibitor based upon preliminary enzymology studies. Gossylic lactone in DMSO was added to monolayers (70% confluent) of TSE cells in Costar 96-well flasks with complete media. Control cells received DMSO or buffer. Cells were incubated for 24hrs, after which CellTiter-96-assay reagent (Promega) was added and the soluble formazan produced by metabolically active cells was read at 490nm. The DMSO had no inhibitory effect. As a reference, HeLa cells, a non-breast cancer cell line, were treated with gossylic lactone. There was no toxicity up to 50 micromolar gossylic lactone.



Available online at www.sciencedirect.com
SCIENCE DIRECT
Chemico-Biological Interactions
ELSEVIER
Chemico-Biological Interactions 143–144 (2003) 481–491
www.elsevier.com/locate/chembioint

17-β-Hydroxysteroid dehydrogenase type 1: computational design of active site inhibitors targeted to the Rossmann fold

William M. Brown [a], Louis E. Metzger, IV [a], Jeremy P. Barlow [b], Lucy A. Hunsaker [a], Lorraine M. Deck [b], Robert E. Royer [a], David L. Vander Jagt [a,*]

[a] *Department of Biochemistry and Molecular Biology, University of New Mexico School of Medicine, Albuquerque, NM 87131, USA*
[b] *Department of Chemistry, University of New Mexico, Albuquerque, NM 87131, USA*

Abstract

17-β-Hydroxysteroid dehydrogenase type 1 (17βHSD1), also called estradiol dehydrogenase, catalyzes the NADPH-dependent reduction of the weak estrogen, estrone, into the more potent estrogen, 17-β-estradiol. 17βHSD1 is an attractive drug target in hormone-sensitive breast cancer. Past efforts to develop selective inhibitors of 17βHSD1 have focused on design of substrate analogs. It is challenging to develop steroid analogs that are devoid of any undesired biological activity. 17βHSD1 is a member of the short-chain dehydrogenase/reductase (SDR) superfamily that includes many hydroxysteroid dehydrogenases. Members of the SDR family bind NAD(P)(H) in a motif that is a modified Rossmann fold. We demonstrated previously that the Rossmann folds of classical dehydrogenases can be selectively inhibited by derivatives and analogs of the natural product gossypol. In this study, we have addressed the question whether the modified Rossmann fold in 17βHSD1 is a target for identification of lead compounds for structure-based drug design. 17βHSD1 was purified from human placenta. 17βHSD1 is inhibited by derivatives of gossypol with dissociation constants as low as 2 μM. Inhibition is competitive with the binding of cofactor. Molecular modeling studies using the published coordinates of human 17βHSD1 suggest that these inhibitors occupy the modified Rossmann fold at the nicotinamide end of the dinucleotide-binding site, extending towards the substrate site. A computational approach was used to design potential new inhibitors of 17βHSD1. The results suggest not only that derivatives of gossypol represent attractive lead compounds for structure-based drug design but also suggest that appropriate incorporation of a substrate analog into the design of these Rossmann fold inhibitors may provide pan-active site inhibitors that span the cofactor and substrate site, potentially offering specificity and increased potency.

Keywords: 17-β-Hydroxysteroid dehydrogenases type 1; Gossypol derivatives; Inhibition; Rossmann fold; Breast cancer

1. Introduction

17β-Estradiol (E_2), the most potent of human estrogens, is known to stimulate the growth of breast cancer cells [1]. In addition, a large fraction of breast tumors are hormone-sensitive. E_2 functions at the nuclear level through interaction with

* Corresponding author. Tel.: +1-505-272-5781; fax: +1-505-272-3518
E-mail address: dlvanderjagt@salud.unm.edu (D.L. Vander Jagt).

0009-2797/02/$ - see front matter
PII: S0009-2797(02)00184-9

the estrogen receptor, leading to subsequent regulation of a battery of genes that control the proliferation of mammary epithelial cells [2]. Consequently, interfering with the mitogenic activities of E_2, either through blocking its production or by inhibiting its receptor interaction, has become a major goal. Attempts to block E_2-receptor interactions have led to the design of inhibitors that are steroid analogs [3]. It is a challenge, however, to create analogs that exhibit selective action against the estrogen receptor, thereby eliminating undesirable biological activities outside this pathway. Therefore, limiting E_2 production may prove to be a more attractive approach to the design of new therapeutics for breast cancer.

E_2 is synthesized locally in peripheral targets from its inactive precursor dehydroepiandrosterone (DHEA) or its sulfate derivative (DHEA-S). This local control of active hormone levels is unique to man and a few primates, and has been termed "intracrinology", distinguishing it from the process by which active hormone is taken from the circulation or extracellular space [4]. In order to synthesize E_2, estrone (E_1) must be produced from DHEA(-S), whether it be in breast epithelia or other tissues. The final reaction, occurring in breast epithelia, reduces the weak estrogen E_1 to the active estrogen E_2. Inhibition of this reaction catalyzed by 17-β-hydroxysteroid dehydrogenase type 1 (17βHSD1) provides a method to lower E_2 production in the target tissue.

17βHSD1 is a member of the short-chain dehydrogenase/reductase (SDR) family. A number of the members of this family utilize nicotinamide adenine dinucleotides (NAD(P)(H)) as cofactors for steroid reduction or oxidation reactions. SDR proteins bind NAD(P)(H) in a motif known as the Rossmann fold, which is the cofactor-binding site in the majority of dehydrogenases [5]. The 17βHSD1 reaction is reversible and dependent on the type of cofactor (NAD(H) or NADP(H)) [6]. In vivo, however, the enzyme acts primarily as a steroid-keto reductase [7], maintaining intracellular levels of E_2. In this reaction, the pro-S hydride from the reduced nicotinamide ring is transferred to the C17 carbonyl of E_1 to form the more potent E_2 [8]. The bisubstrate reaction is reported to occur via a random mechanism [9], providing two sites that can be targeted for inhibition, i.e., the E_2-binding site and the Rossmann fold.

We have previously demonstrated that the natural product gossypol, a polyphenolic binaphthyl isolated from cottonseed, inhibits all isozymes of human lactate dehydrogenase (LDH), which also contain the Rossmann fold. Several derivatives of gossypol, along with many analogs, have been synthesized. These compounds exhibit a range of selectivities for human LDHs, with inhibition constants as low as 30 nM [10–12]. Inhibition by these compounds is consistently competitive with the binding of NADH. These data, along with the structural conservation of the Rossmann fold across many oxidoreductase enzymes, suggest that these compounds may represent lead structures for design of inhibitors of dehydrogenases that possess a Rossmann fold. In this study, we evaluated gossypol, gossypol derivatives, and gossypol analogs as inhibitors of human 17βHSD1. In addition, computational approaches were used to model 17βHSD1-ligand interactions, and to suggest a further direction for the design of new inhibitors.

2. Materials and methods

2.1. Synthesis of gossypol analogs and derivatives

Derivatives and analogs of gossypol were prepared as described previously [10,13–15].

2.2. Protein purification

The protocol for purification of 17βHSD1 was a modification of that reported by Yang et al. [16]. Fresh human placenta, 250 g, was cubed and homogenized. The 100,000 × *g* supernatant fraction was purified on a Blue Sepharose CL-6B column. The eluent was concentrated by pressure filtration, desalted on a PD-10 column and chromatofocused. Enzyme purity was examined using SDS-PAGE electrophoresis on a 20% separating gel with a 5% stacking gel.

2.3. Enzyme assay

17βHSD1 activity in the direction of oxidation of estradiol to estrone was measured in 1 ml total volume of 0.1 M sodium bicarbonate buffer pH 9.2, 25 μM estradiol, and 0.5 mM β-NAD. Enzyme activity was determined by following changes in NAD concentration at 340 nm, $\varepsilon = 6.2\ mM^{-1}\ cm^{-1}$.

2.4. Enzyme kinetic studies

Initial velocity studies were conducted in the buffers described above at 25 °C. Michaelis constants for substrates and cofactors and k_{cat} values were determined by nonlinear regression analysis of the initial rate data using the ENZFITTER program (Elsevier-Biosoft).

2.5. Flexible docking

Flexible docking of the inhibitors to human 17βHSD1 was performed using the AutoDock 3.0 software suite from Scripps Research Institute [17]. The crystal structure of 17βHSD1 (1A27.pdb) was modified to accommodate the docking [18]. The coordinates of polar hydrogens were added as predicted by Sybyl 6.6 using torsional minimization. Partial charges were assigned from united Kollman dictionary charges and all substrate and ordered water atoms were removed. Inhibitor structure was predicted from GA conformational search followed by BFGS minimization in Sybyl. Inhibitor partial charges were assigned according to the Gasteiger–Huckel method.

2.6. Active site analysis

In an effort to aid rational design of improved inhibitors based on theoretical docking studies, an algorithm was implemented in C++ to evaluate free space within the enzyme around a docked inhibitor. For each atom in a docked inhibitor, points are evaluated around a sphere of radius 1.5 Å. If a pseudo-atom placed at this point experiences no steric clash with protein or inhibitor atoms, a graphical dot is placed at this *X*-, *Y*-, and *Z*-coordinate and a new sphere of points is evaluated around this dot. Thus, graphical dots will only be seen at positions in the inhibitor where an atom with van der Waals radius of 1.5 Å will fit, and only at positions within average bond lengths of inhibitor atoms. Recursive analysis of points at bonding distances from the inhibitor has an advantage over grid-based evaluation methods in that only free space continuous with inhibitor atoms will be shown. Additionally, this allows direct comparison between active sites across enzymes, an important consideration in drug design that is not possible with other analysis programs. Each graphical point is colored according to its electrostatic potential, calculated using the distance-dependent dielectric of Mehler and Solmajer [19] to model bulk solvent effects. The algorithm provides a visual identification of areas around the inhibitor that might be utilized to increase selectivity and binding energy.

3. Results

3.1. Compound screening

Seven gossypol-related compounds were screened against 17βHSD1 at pH 9.2. All inhibitors were tested at a concentration of 25 μM with 0.5 mM NAD and 25 μM E_2. Addition of gossypol resulted in only a slight reduction in enzyme activity (Fig. 1). Four gossypol derivatives, in which the aldehyde functional group is modified, were tested. The peri-acylated nitriles, gossylic nitrile 1,1′-diacetate (GNDA) and gossylic nitrile 1,1′-divalerate (GNDV), represent compounds in which the aldehyde group is converted to a nitrile, and the peri-hydroxyl of gossypol is derivatized. GNDA showed a 13% reduction in activity, while GNDV showed no reduction in activity at these concentrations. Gossylic iminolactone (GIL) and gossylic lactone (GL) were the most promising compounds, producing 80–90% reduction in enzyme activity. Three gossypol analogs (2,3-dihydroxynaphthoic acids with different substituents at the 4- and 7-positions) exhibited little effect against 17βHSD1.

Inhibition constants for GL and GIL were determined from initial velocities at 3 μM inhibitor

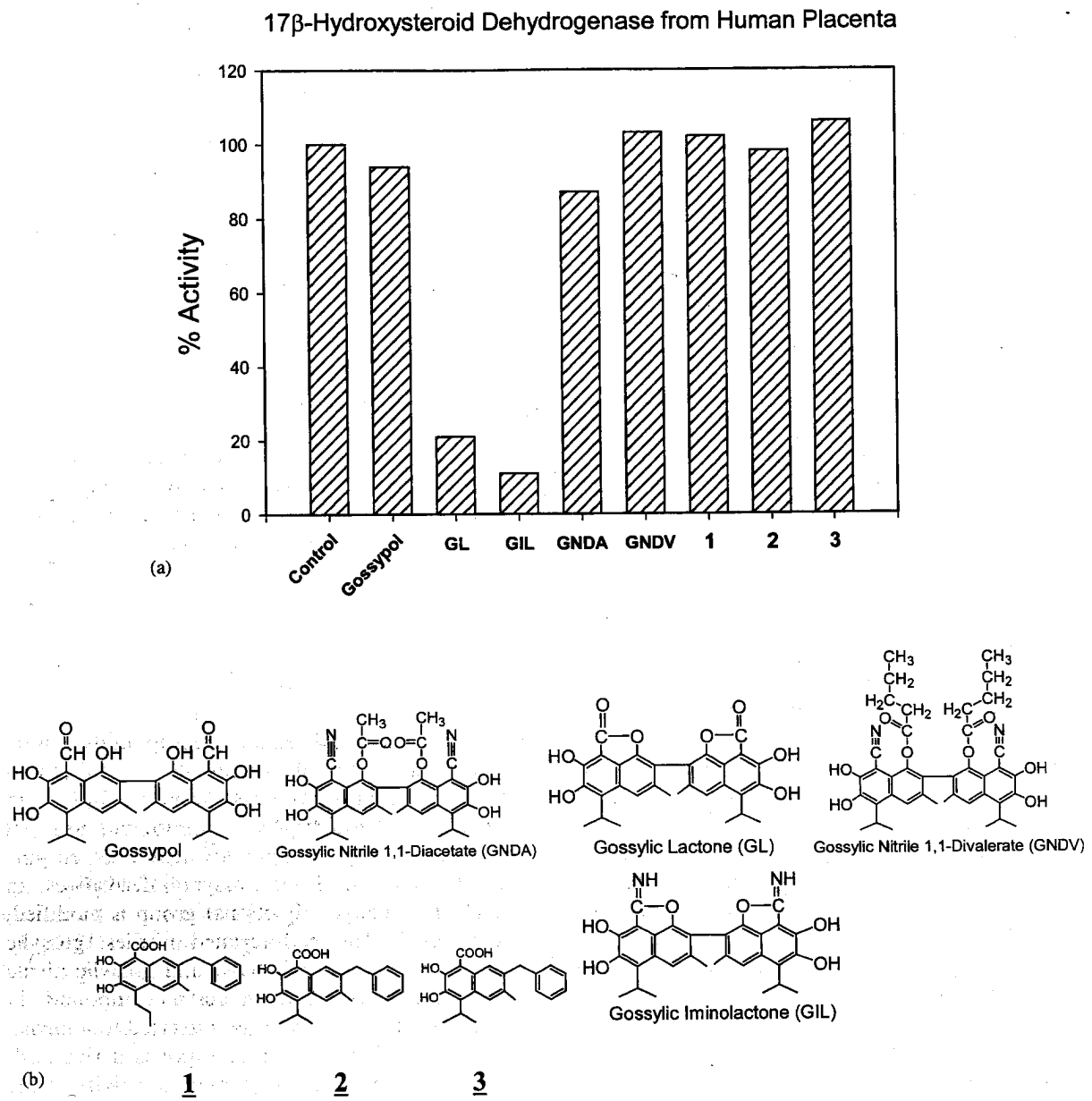


Fig. 1. Inhibition of 17βHSD1 by gossypol, gossypol derivatives and analogs of gossypol in the 2,3-dihydroxynaphthoic acid family.

concentrations. Both compounds exhibited competitive inhibition with respect to NAD. The inhibition constants for GL and GIL were determined to be 2.2 and 4.3 μM, respectively. A representative Lineweaver–Burk plot for inhibition of 17βHSD1 by GL is shown in Fig. 2.

3.2. 17βHSD1-inhibitor complex prediction

In order to predict the binding modes of GL and GIL in complex with 17βHSD1, flexible docking studies were performed. The crystal structure of 17βHSD1 in complex with estradiol and $NADP^+$

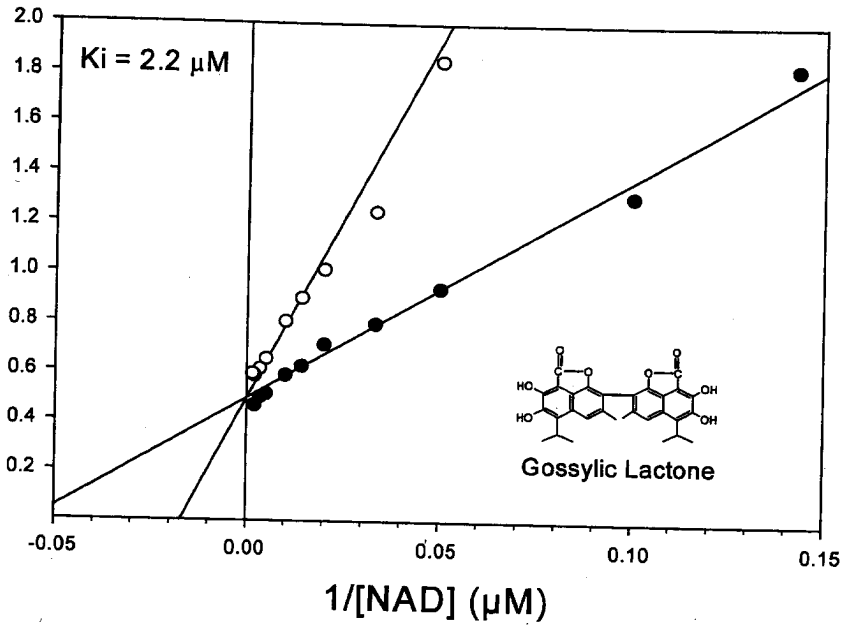


Fig. 2. Inhibition of 17βHSD1 by gossylic lactone is competitive with the binding of cofactor.

(1.9 Å resolution, 0.21 *R*-value) as determined by Mazza et al. [18] was used for the studies. GL docked in an orientation completely within the cofactor site, lying in the region towards the nicotinamide residue-binding area (Fig. 3). In this orientation, the compound exhibits important hydrogen bonds with Y155, S142, G141, K159, L93, G92, K195, and R37 (Fig. 4). GIL was also docked, resulting in orientations similar to that of GL.

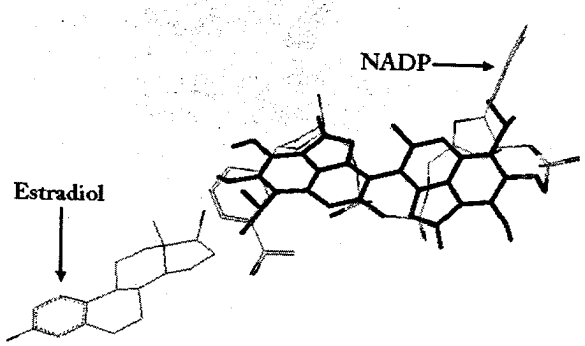


Fig. 3. Modeled structure of 17βHSD1 complexed with gossylic lactone.

3.3. Active site analysis

Recursive analysis of space around the docked inhibitors that might be used to increase the selectivity or binding affinity of the compounds was performed using a new algorithm implemented in C++. The algorithm is intended for lead optimization purposes, and provides a graphical representation of the steric and electrostatic properties of the binding pocket space surrounding an inhibitor. In the case of GL, the substrate-binding pocket is continuous with groups at the 5-(isopropyl) and 6-(hydroxyl) positions of GL (Fig. 5). Most of the inhibitor exhibits tight contacts with the binding site, aside from the 5′- and 6′-positions on the other side of the molecule that lead into the adenine-binding region. The results suggest that modifications of the 5- or 6-positions with substituents that take advantage of the E_2-binding region may be useful for acquiring additional binding affinity and selectivity. The results from this study were used to suggest a scaffold structure for further design of inhibitors (discussed in the following section).

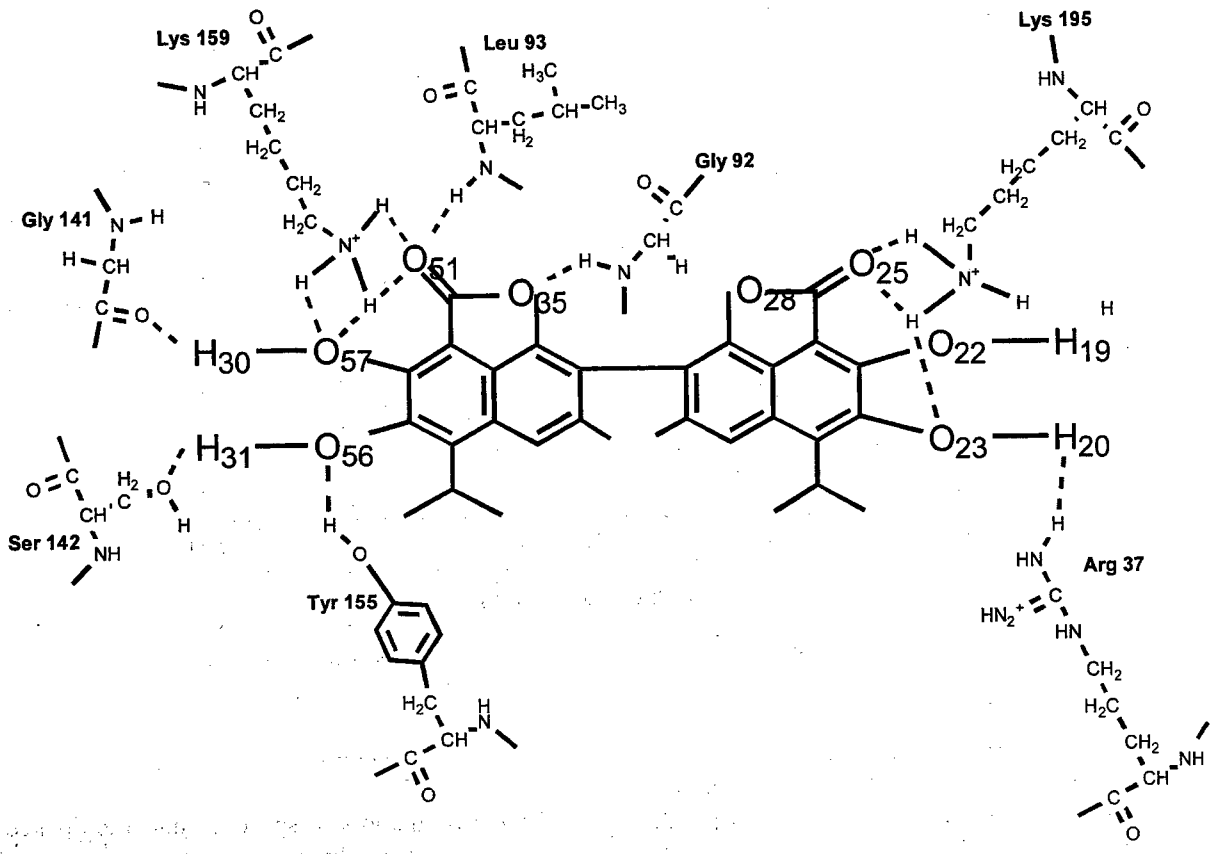


Fig. 4. Hydrogen bonding interactions between 17βHSD1 and gossylic lactone from modeling.

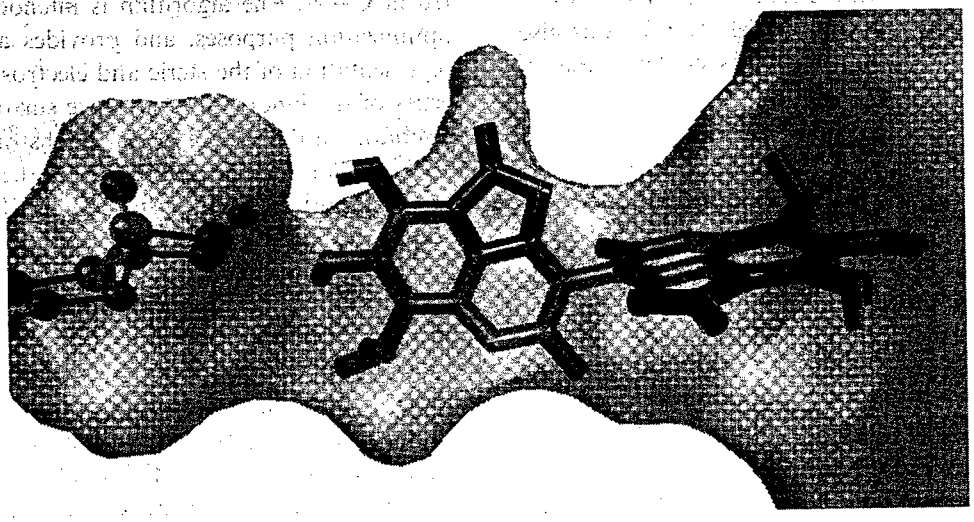


Fig. 5. Active site analysis of 17βHSD1 from the docked complex of gossylic lactone focused on the nicotinamide end of the Rossmann fold extending into the substrate site. For reference, the C and D rings of estradiol are shown.

4. Discussion

4.1. Gossypol-related compounds as lead structures for the inhibition of Rossmann folds

Early characterization of the dinucleotide-binding sites of a series of dehydrogenases was reported by Rossmann et al. [20], who defined the structural conservation in the cofactor-binding sites of LDH, alcohol dehydrogenase, glyceraldehyde-3P dehydrogenase, and malate dehydrogenase. Subsequent characterization of the structures of other dinucleotide-binding proteins resulted in an extended definition of the Rossmann fold by Wierenga et al. [21] and Bellamacina [22]. The "classical" Rossmann fold is defined for proteins resembling LDH. In these two-domain proteins, it is the carboxy-terminal domain that is responsible for cofactor binding. The minimum core topology of the classical Rossmann fold includes a βαβαβ unit (two α-helices packed on one side of a three-stranded parallel β-sheet) associated with a fourth β-strand. The fourth strand usually constitutes the first part of a second βαβαβ unit, related to the first by a roughly twofold symmetry. Associated with the secondary structures is a glycine-rich consensus sequence GXGXXG necessary for the tight packing of secondary structure elements. Additional primary structure conservation of six small hydrophobic residues along with an R or K is also present. The ADP part of NAD(P) binds to the core unit; the nicotinamide ring resides in a crevice between the fourth β-strand (i.e., the first β-strand of the second unit) and the remainder of the second βαβαβ unit. Recent evolutionary studies suggest that, although the dinucleotide-binding site likely evolved from gene duplication, the two sites evolved separately, with the N-terminal unit showing less variation than the C-terminal unit [23].

From these studies, it became clear that dehydrogenases from different families and superfamilies can have structural homology at the NAD(P)-binding domains, even in the absence of any significant sequence homology (aside from a few conserved residues). However, not all families containing NAD(P)-binding proteins use the classical Rossmann fold for binding. Aldehyde dehydrogenases represent a newly recognized superfamily of proteins containing a modified Rossmann fold in which the signature sequence (GXGXXG) is missing. These three-domain structures bind NAD(P) quite differently than the classic Rossmann dehydrogenases, utilizing five rather than six β-strands [24]. Dehydrogenases/reductases in the aldo-keto reductase superfamily do not contain any motif resembling the Rossmann fold, but rather use a single-domain $(\alpha/\beta)_8$ barrel (TIM barrel) to bind cofactor [25].

The SDR family, a large family that includes prostaglandin dehydrogenase and a number of important hydroxysteroid dehydrogenases, represents a family with a slightly modified Rossmann fold [5]. These enzymes are single-domain proteins where the Rossmann fold is located near the N-terminus and contains seven or eight β-strands rather than the normal six strands. The signature motif, GXXXGXG, differs somewhat from that of the classic Rossmann fold but appears to have the same function in cofactor binding. In addition, there is a conserved YXXXK sequence that interacts with the Rossmann fold.

We have previously demonstrated that gossypol-related compounds competitively inhibit cofactor binding at the classical Rossmann fold of both human and parasitic LDHs [10–12]. In this study, we addressed the question of whether or not these compounds will be effective inhibitors of 17βHSD1, an enzyme from the SDR family that contains a modified Rossmann fold. It may seem inappropriate to propose that these inhibitors will also be effective against 17βHSD1 based on secondary structure similarities alone. However, the orientation of NAD(P) in Rossmann fold binding domains consists of an extended conformation that is remarkably similar across enzymes [22]. This overlap of cofactor conformations suggests a conservation of interaction sites at the tertiary level that would be necessary to maintain the binding mode, even if these interactions do not result from primary structure conservation. In the case of 17βHSD1, the orientation of the cofactor differs from that of LDHs in that the nicotinamide ring is in the syn rather than anti-conformation, leading to transfer of the pro-S (B-face) rather than pro-R hydride [8,20]. Despite this difference,

in our modeling studies, the $NADP^+$ from the crystal structure of 17βHSD1 [18] results in only a 2.46 Å RMSD when superimposed with the NAD^+ from the crystal structure of a parasitic LDH [26]. Most of this difference comes from the nicotinamide ring orientation; the conformations of the remaining portions of the cofactors are nearly identical, resulting in less than a 1.3 Å RMSD when the cofactors are superimposed without the nicotinamide ring present.

Despite its deviations from the classical Rossmann fold, we have now shown that several gossypol-related compounds exhibit competitive inhibition of cofactor binding in 17βHSD1. The data further support the concept of using gossypol-related compounds as lead structures for the development of inhibitors targeted to Rossmann folds. However, important to the concept of a lead compound is the ability to derive from it a selective inhibitor. The argument that structural conservation provides the basis for compounds that are leads for several drug targets raises a concern about selective inhibition. How can selective inhibitors be developed for a conserved structural motif that is present in so many types of enzymes? The same question arose as an argument against kinase inhibitors targeted to the kinase ATP-binding site, however, and kinase inhibitors are now in clinical testing [27]. Human LDHs (LDH-A_4, B_4, and C_4) are highly homologous with 84–89% similarity and 69–75% identity. The amino acids that comprise the Rossmann fold for these isozymes are nearly identical. Nevertheless, for gossypol analogs in the 2,3-dihydroxynaphthoic acid family, greater than 200-fold selectivity was observed when substituents at the 4- and 7-positions were varied [12]. Additionally, in this study we show that some of these promising analogs for human LDH inhibition show little inhibition of 17βHSD1, providing further evidence that selective inhibitors that target the Rossmann fold can be developed.

4.2. GL as a lead compound for inhibition of 17βHSD1

Tumor cell metabolism is considered to be an important factor in the pathogenesis and development of various sex steroid-dependent neoplasms, via the synthesis of estrogens. In the majority of human breast cancers, estrogens have been shown to contribute to neoplastic progression, and some breast cancers are unable to sustain growth in the absence of estrogen [28]. The biologically active estrogen E_2 is synthesized through a reaction catalyzed by 17βHSD1. Increased expression of 17βHSD1 has been observed in all of the clinical stages of breast carcinoma suggested to represent neoplastic progression, and the enzyme is thought to be a significant factor in early progression of the disease [29]. It can, therefore, be seen that targeting 17βHSD1 is an important consideration in the search for breast cancer therapeutics.

17βHSD1 is an SDR enzyme with unique characteristics associated with its substrate specificity and catalytic function [8]. With 327 residues, 17βHSD1 is one of the largest SDR enzymes. The SDR YXXXK motif is present, and six of the seven residues often expressed by SDR at coenzyme-binding sites are observed. Additionally, the basic residue normally located in the consensus sequence of the dinucleotide-binding motif (GXXXGXG) is replaced with S12. The C-terminal substrate-binding site of SDR proteins is the most variable, and, in the case of 17βHSD1, the presence of H221 and E282 in the steroid-binding cleft provide for substrate specificity via a bifurcated hydrogen bond with the steroid 3-hydroxyl group [18]. Catalytically relevant steroid/protein interactions (at O17 of estradiol) are maintained by Y155 and S142. A flexible loop at residues 191–199 is stabilized by cofactor binding, and appears to protect NAD(P)(H) from solvent.

We have demonstrated here that GL and GIL exhibit micromolar inhibition constants with competitive binding against NAD(P)(H). Docking studies suggest that the inhibitors reside primarily in the nicotinamide-binding region. This is consistent with modeling and fluorescence-quenching studies performed with gossypol-related compounds and LDH from *Plasmodium falciparum* (unpublished data). The docked orientation of GL is held by hydrogen bonds to SDR signature motif residues Y155 and K159 (YXXXK), and to an additional catalytic residue S142. Additional hydrogen bonds on the same side of the inhibitor

involve L93 and G141. The opposing side of the inhibitor participates in hydrogen bonds with R37 and K195. These residues stabilize the 17βHSD1 flexible loop through charge compensation via salt bridges with the dinucleotide 2′-phosphate [18]. It is, therefore, possible that inhibitor interactions with these residues play a role in stabilizing the flexible loop in the absence of cofactor and would explain the preference of 17βHSD1 for GL and GIL compared with dihydroxynaphthoic acid analogs (Fig. 1). Active-site analysis suggests that in this orientation, substituent modification would only be plausible at the 5-, 6-, 5′-, and 6′-positions of the naphthalene rings. The 6-position is involved in hydrogen bonding with 17βHSD1, however, the isopropyl at the 5-position heads away from cofactor-binding interactions towards the catalytic site. This identifies an important area for substituent modification that might be necessary for increasing binding and selectivity. The position of the inhibitor within the nicotinamide-binding region, which is the most variable among Rossmann folds, and the position of the 5-substituent suggests that GL is a promising lead compound from which selective inhibitors can be derived.

4.3. Pan-active site inhibitors of 17βHSD1

As with any compound, it is difficult to assess the presence of any unwanted biological effects that may result from inhibition before clinical trials. These effects may result from unforeseen influences of target pathway inhibition or from nonspecific inhibitor binding. The limited number of available protein structures prevents any suitable attempt for computational prediction of the latter. The cofactor for bisubstrate reactions is usually not unique to one enzyme, and the differences in substrate specificities across enzymes can be subtle. This makes the idea of pan-active site inhibition, i.e., inhibitor competition for both substrate and cofactor sites, very attractive. Ideally, substrate specificity of one portion of an inhibitor along with cofactor specificity from another portion would limit compound binding to enzymes catalyzing only one type of reaction. This idea of pan-active site inhibition is especially attractive for inhibitors targeted to a motif such as the Rossmann fold that is present in such a large variety of proteins as a method to improve selectivity.

Docking studies suggest that substitution at the 5-(isopropyl) position of GL can be utilized to generate pan-active site inhibitors for 17βHSD1. Such inhibitors would be somewhat large, however, and could result in difficulties with active-site accessibility and compound synthesis. We, therefore, made modifications to hemigossylic lactone (hGL), which represents only one-half of the symmetric compound, for computational studies. Use of hGL inhibitors may be of additional importance in that the compounds do not exhibit the atropoisomerism (isomers that result from hindered rotation about the binaphthal bond) seen in GL. Consequently, there is no concern about the differences in activity resulting from racemic mixtures of inhibitor. Additionally, docking studies suggest that binding interactions with R37 and K195 may still be obtained via small substituents at the 2-position of hGL. Substitution at the 5-position with a butylene attached to a substrate mimetic (C and D rings of estradiol), results in a compound that docks in a pan-active orientation within the active site (Fig. 6). The docked energy of the compound is nearly doubled with respect to that of GL. Therefore, the compound provides a promising structure from which synthetically accessible compounds may be derived that utilize the substrate-binding site for increased binding affinity and selectivity.

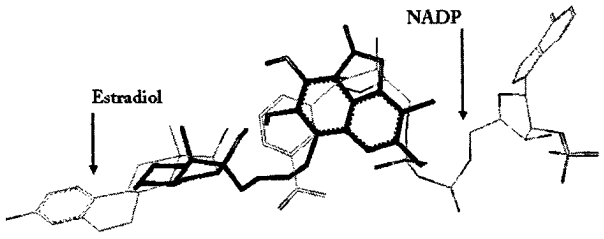


Fig. 6. Modeled structure of 17βHSD1 complexed with a designed pan-active site inhibitor corresponding to hGL except that a substrate analog (corresponding to the C and D rings of estradiol) replaced the isopropyl group in the 5-position.

Acknowledgements

This work was supported by US Army/DOD Breast Cancer Program grants DAMD17-00-0372 (DLVJ) and predoctoral awards DAMD17-00-1-0368 (WMB) and DAMD17-00-0369 (JPB).

References

[1] H. Peltoketo, P. Vihko, R. Vihko, Regulation of estrogen action: role of 17 beta-hydroxysteroid dehydrogenases, Vit. Horm. 55 (1999) 353–398.

[2] T.M. Penning, Molecular endocrinology of hydroxysteroid dehydrogenases, Endocr. Rev. 18 (1997) 281–305.

[3] M.R. Tremblay, D. Poirier, Overview of a rational approach to design type I 17beta-hydroxysteroid dehydrogenase inhibitors without estrogenic activity: chemical synthesis and biological evaluation, J. Steroid Biochem. Mol. Biol. 66 (1998) 179–191.

[4] F. Labrie, Intracrinology, Mol. Cell. Endocrinol. 78 (1991) C113–C118.

[5] H. Jornvall, B. Persson, M. Krook, S. Atrian, R. Gonzalez-Duarte, J. Jeffery, D. Ghosh, Short-chain dehydrogenases/reductases (SDR), Biochemistry 34 (1995) 6003–6013.

[6] J.Z. Jin, S.X. Lin, Human estrogenic 17beta-hydroxysteroid dehydrogenase: predominance of estrone reduction and its induction by NADPH, Biochem. Biophys. Res. Commun. 259 (1999) 489–493.

[7] V. Luu-The, Y. Zhang, D. Poirier, F. Labrie, Characteristics of human types 1, 2 and 3 17 beta-hydroxysteroid dehydrogenase activities: oxidation/reduction and inhibition, J. Steroid Biochem. Mol. Biol. 55 (1995) 581–587.

[8] R. Breton, D. Housset, C. Mazza, J.C. Fontecilla-Camps, The structure of a complex of human 17beta-hydroxysteroid dehydrogenase with estradiol and NADP+ identifies two principal targets for the design of inhibitors, Structure 4 (1996) 905–915.

[9] G. Betz, Reaction mechanism of 17-beta-estradiol dehydrogenase determined by equilibrium rate exchange, J. Biol. Chem. 246 (1971) 2063–2068.

[10] L.M. Deck, R.E. Royer, B.B. Chamblee, V.M. Hernandez, R.R. Malone, J.E. Torres, L.A. Hunsaker, R.C. Piper, M.T. Makler, D.L. Vander Jagt, Selective inhibitors of human lactate dehydrogenases and lactate dehydrogenase from the malarial parasite *Plasmodium falciparum*, J. Med. Chem. 41 (1998) 3879–3887.

[11] D.L. Vander Jagt, L.M. Deck, R.E. Royer, Gossypol: prototype of inhibitors targeted to dinucleotide folds, Curr. Med. Chem. 7 (2000) 479–498.

[12] Y. Yu, J.A. Deck, L.A. Hunsaker, L.M. Deck, R.E. Royer, E. Goldberg, D.L. Vander Jagt, Selective active site inhibitors of human lactate dehydrogenases A4, B4, and C4, Biochem. Pharmacol. 62 (2001) 81–89.

[13] R.E. Royer, L.M. Deck, N.M. Campos, L.A. Hunsaker, D.L. Vander Jagt, Biologically active derivatives of gossypol: synthesis and antimalarial activities of peri-acylated gossylic nitriles, J. Med. Chem. 29 (1986) 1799–1801.

[14] L.M. Deck, D.L. Vander Jagt, R.E. Royer, Gossypol and derivatives: a new class of aldose reductase inhibitors, J. Med. Chem. 34 (1991) 3301–3305.

[15] R.E. Royer, L.M. Deck, T.J. Vander Jagt, F.J. Martinez, R.G. Mills, S.A. Young, D.L. Vander Jagt, Synthesis and anti-HIV activity of 1,1′-dideoxygossypol and related compounds, J. Med. Chem. 38 (1995) 2427–2432.

[16] F. Yang, D.W. Zhu, J.Y. Wang, S.X. Lin, Rapid purification yielding highly active 17 beta-hydroxysteroid dehydrogenase: application of hydrophobic interaction and affinity fast protein liquid chromatography, J. Chromatogr. 582 (1992) 71–76.

[17] G.M. Morris, D.S. Goodsell, R.S. Halliday, R. Huey, W.E. Hart, R.K. Belew, A.J. Olson, Automated docking using a Lamarckian genetic algorithm and empirical binding free energy function, J. Comp. Chem. 19 (1998) 1639–1662.

[18] C. Mazza, R. Breton, D. Housset, J.C. Fontecilla-Camps, Unusual charge stabilization of NADP+ in 17beta-hydroxysteroid dehydrogenase, J. Biol. Chem. 273 (1998) 8145–8152.

[19] E.L. Mehler, T. Solmajer, Electrostatic effects in proteins: comparison of dielectric and charge models, Protein Eng. 4 (1991) 903–910.

[20] M.G. Rossmann, A. Liljas, C.I. Branden, L.J. Banaszak, in: P. Boyer (Ed.), The Enzymes, Academic Press, New York, 1975, pp. 61–102.

[21] R.K. Wierenga, M.C.H. De Maeyer, W.G.J. Hol, Interaction of pyrophosphate moieties with alpha-helixes in dinucleotide binding proteins, Biochemistry 24 (1985) 1346–1357.

[22] C.R. Bellamacina, The nicotinamide dinucleotide binding motif: a comparison of nucleotide binding proteins, FASEB J. 10 (1996) 1257–1269.

[23] O. Carugo, P. Argos, NADP-dependent enzymes. II. Evolution of the mono- and dinucleotide binding domains, Proteins 28 (1997) 29–40.

[24] S.Y. Liu, Z.J. Rose, Y.J. Chung, C.D. Hsiao, W.R. Chang, I. Kuo, J. Perozich, R. Lindahl, J. Hempel, B.C. Wang, The first structure of an aldehyde dehydrogenase reveals novel interactions between NAD and the Rossmann fold, Nat. Struct. Biol. 4 (1997) 317–326.

[25] D.K. Wilson, K.M. Bohren, K.H. Gabbay, F.A. Quiocho, An unlikely sugar substrate site in the 1.65 Å structure of the human aldose reductase holoenzyme implicated in diabetic complications, Science 257 (1992) 81–84.

[26] C.R. Dunn, M.J. Banfield, J.J. Barker, C.W. Higham, K.M. Moreton, D. Turgut-Balik, R.L. Brady, J.J. Holbrook, The structure of lactate dehydrogenase from *Plasmodium falciparum* reveals a new target for antimalarial design, Nat. Struct. Biol. 3 (1996) 912–915.

[27] C.L. Arteaga, F. Khuri, G. Krystal, S. Sebti, Overview of rationale and clinical trials with signal transduction inhibitors in lung cancer, Semin. Oncol. 29 (2002) 15–26.
[28] I.C. Henderson, G.P. Canellos, Cancer of the breast: the past decade (first of two parts), N. Engl. J. Med. 302 (1980) 17–30.
[29] H. Sasano, T. Suzuki, J. Takeyama, H. Utsunomiya, K. Ito, N. Ariga, T. Moriya, 17-beta-Hydroxysteroid dehydrogenase in human breast and endometrial carcinoma: a new development in intracrinology, Oncology 59 (Suppl. 1) (2000) 5–12.

New Approach for Characterization of Binding-Site Search Space

W. Michael Brown, David L. Vander Jagt
Department of Biochemistry and Molecular Biology, University of New Mexico School of Medicine, Albuquerque, New Mexico 87131

ABSTRACT: A new approach for defining the Cartesian spatial boundaries of binding pockets is presented. The method involves calculation of a *macromolecule encapsulating surface* (MES) that separates binding pocket volume from outside space. The surface provides means for identification of binding sites and calculation of their volume. Additionally, the MES can be used to limit the search space for ligand docking and *de novo* design algorithms via identification of accessible atoms within the binding pocket or limitation of translation ranges to binding pocket space. The approach has been shown to be efficacious based on testing with 50 enzyme-ligand complexes for which the binding pockets are known. Additionally, we have modified the flexible docking program AutoDock 3.0 to incorporate MES boundaries using an energetic term. The results show increased efficiency of the genetic algorithm for ligand docking characterized by a larger percentage of successful runs and a decrease in required run times. MES incorporation also facilitates search of an entire enzyme for ligand docking, without the requirement of a predetermined binding pocket location.

Keywords: binding pockets; active sites; ligand docking; *de novo* design; genetic algorithm

Correspondence to W. M. Brown; e-mail: mbrown@unm.edu

Grant Sponsor: US Army/DOD Breast Cancer Program, grant numbers DAMD17-00-0368, DAMD17-00-0372

Introduction

Biological metabolism is largely regulated via enzymatic catalysis of biochemical reactions. In this process, a substrate interacts with a macromolecule enzyme whose role is to decrease energy barriers in order to facilitate chemical reaction under physiologic conditions. Most often, this substrate ligand "binds" to a concave region (termed binding pocket here) within the enzyme's surface. The concave nature of the binding pocket within an enzyme would seem to serve several purposes – it provides for interatomic interactions sufficient to facilitate favorable enthaplic energies, it provides for enzyme-substrate specificity, it orients ligand atoms needing chemical modification with active-site atoms or other ligands, and it can lower the energetic cost of assuming ligand conformations that favor reaction (transition-state conformations). Therefore, the binding pocket structure is of intense biological interest for those who wish to understand or possibly alter biochemical mechanisms.

There are many computational tools that focus on binding pocket structure in their analyses. These include ligand docking software, *de novo* design/lead optimization software, and software rendering abstract graphical representations of ligand-binding pocket interactions. Ligand docking tools aim to predict if a ligand will bind to a macromolecule, and if so, the binding conformation(s) assumed by the ligand. *De novo* design and lead-optimization programs take the binding pocket (possibly with a lead inhibitor bound) and attempt to build ligands predicted to elicit high binding affinity. Most of these algorithms can be roughly grouped into two categories based on their characterization of search space. *Matching algorithms* identify the binding pocket and the ligand as a set of interaction sites (for examples see references 1-4). These algorithms then solve the problem of matching complementary interaction sites to find the fit that results in the highest score. *Force field-based algorithms,* on the other hand, utilize a potentially infinite search space, performing optimization of an objective function describing binding energy parameterized by ligand orientation (for examples see references 5-8).

Both types of algorithm rely on some limitation of search space in order to achieve acceptable run times. Matching algorithms must choose those macromolecule atoms in the binding pocket that will be utilized for interaction site generation. Force-field based algorithms must decide how to limit Cartesian translation ranges around the active site. The difficulty met in both approaches is that there is no concrete definition of the spatial boundaries that characterize the binding pocket region of the enzyme. The macromolecule atoms composing the binding pocket create a spatial boundary due to the repulsive intermolecular interactions with ligand atoms. However, necessary to the concept of a binding pocket is an opening for ligand access. This opening provides for an infinite space continuous with the binding pocket.

Thus, an ambiguity arises in that it is not easy to decide where binding pocket volume ends and space outside the enzyme begins. This transition has been termed as the "sea level"[9] of the binding pocket and the ambiguity has been described as the "can of worms" problem[10]. Due to these difficulties, most docking and *de novo* design applications limit the search space based on a sphere or grid centered on the binding pocket, or based on atom contacts surrounding a ligand in a predetermined binding mode.

These approaches are advantageous in that they are easy to implement and provide for fast bounds checking. However, they are not without several important drawbacks. First, they require *a priori* knowledge of a ligand binding mode or active site residues. Second, the dimensions of the binding pocket are determined based on a known ligand or guess – not based on the volume composing the binding pocket. Finally,

spheres and grids are likely to include search space that is not relevant to the calculations.

Here, we present a new approach for defining the binding pocket search space. The method involves calculation of a *macromolecule encapsulating surface* (MES). The MES separates volume inside the macromolecule from volume outside. A range of continuous empty volume within the macromolecule represents a binding pocket. This range can be used to limit translations in force-field based algorithms, ligand growth in *de novo* design algorithms, and the accessible atoms evaluated for interaction site generation in matching algorithms. Binding pockets may be selected based on their encapsulation of known ligands, active site residues, or criteria such as volume or other limited shape descriptors.

Additionally, we compare the efficiency of the flexible docking program AutoDock 3.0 when MES boundaries are enforced via pre-calculated energetic terms. The testing is performed on 14 enzyme-ligand complexes using the genetic algorithm (with and without local search).

Methods

MES GENERATION – *Algorithm 1*

The MES is defined as the surface that encapsulates the macromolecule and separates binding pocket volumes from those outside the macromolecule. The observation that concave regions within a protein characterize binding pockets suggests a surface with some restraint such that the surface cannot curve "inside" into binding pocket space. We can then consider the macromolecule to have a minimum surface covering the "outside" of the molecule, and adjust this minimum surface so that the restraint is satisfied.

Using known definitions, this could be accomplished by considering the solvent-accessible surface[11] to represent the minimum surface. We could then create the MES by expanding this surface such that a curvature restraint (based on a differential geometry definition[12]) is satisfied. This creates problems from an algorithmic standpoint, however. The solvent accessible surface is discrete and steps would be necessary to control the direction of surface points during expansion such that large holes would not be generated. Additionally, assessing the curvature of the surface would require estimating second derivative information in a discrete setting.

The fact that we do not want the surface to curve under and into binding pockets allows for a simpler, presumably faster, algorithm in which the surface points are restrained to lie on defined variables in spherical coordinates. Under this restriction and the van der Waals hard sphere approximation[13] for steric considerations, we can define the minimum surface as the set of surface points S defined by the spherical coordinates ρ, θ, and ϕ, with the origin (O) located at the macromolecule center of mass. The minimum surface is parameterized by the shell space (S), which is the smallest distance allowed between the minimum surface and the van der Waals surface[13] of the macromolecule.

The initial surface is the set of points $S_i[\rho_i, \theta_i, \phi_i] \in S$ for each surface point index i, where ρ_i is set to some initial value ($\|\boldsymbol{a}\|$) that allows the points to encompass the macromolecule (see below). The minimum surface can then be calculated as the set of points $S_i[\rho_i, \theta_i, \phi_i]$ where ρ_i is the set minimum of P where

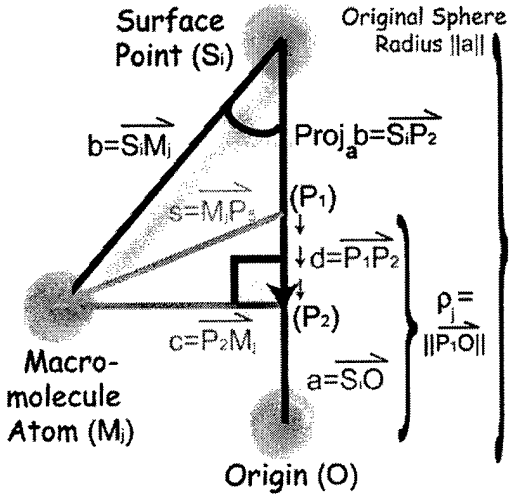
$$\rho_j \in P,\ \rho_j = \|a_j\| - \left(\frac{b_j^T a_j}{\|a_j\|} - \left(\|s_j\|^2 - \|c_j\|^2 \right)^{1/2} \right)$$

$$\|s_j\| = \mathrm{vdW}(M_j) + S$$

$$c_j = b_j - proj_a b_j$$

for each macromolecule atom index j, when $\boldsymbol{b_j}$ represents the vector between $\boldsymbol{S_i}$ and the center of macromolecule atom $\boldsymbol{M_j}$, $\boldsymbol{a_j}$ represents the vector

between the initial surface point S_i and O, and vdW(M_j) represents the van der Waals radius of the atom (figure 1).



Because a curvature restraint has little meaning under the spherical coordinate restriction, a *compression restraint* is proposed that controls the drop of a surface point towards the macromolecule as a function of the distance between the points:

$$f(\rho,\phi,\theta) \le d\rho \le g(\rho,\phi,\theta)$$
$$f(\rho) = -g[\rho - f(\rho)]$$

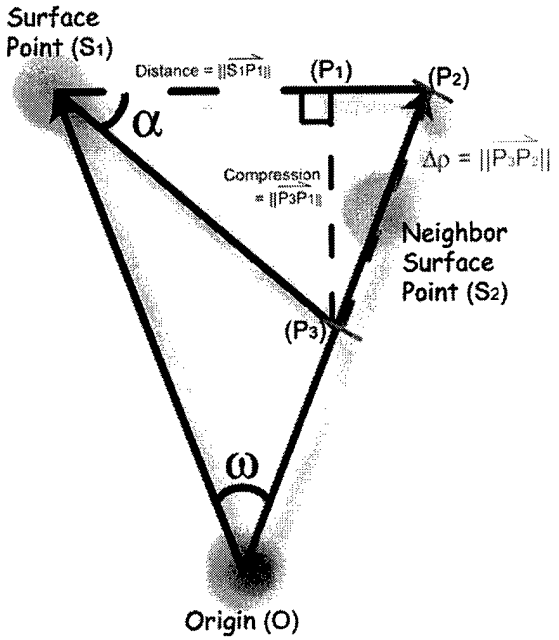
Because the compression restraint should be uniform around a sphere at a given ρ, $d\theta$ (θ_2-θ_1) and dϕ (ϕ_2-ϕ_1) are expressed in terms of ω which is the angle between the rays at $[\theta_2,\phi_2]$ and $[\theta_1,\phi_1]$:

$$\omega(d\theta,d\phi) \equiv 2 \bullet \sin^{-1}\left(\frac{2-\cos(d\theta)-\cos(d\phi)}{2}\right)^{1/2}$$

Then, the compression restraint, as implemented in discrete form, can be defined in spherical coordinates as

$$\Delta\rho(\Delta\theta,\Delta\phi) \ge \frac{2 \bullet \rho \sin(\frac{\omega}{2}) \sin\alpha}{\cos(\alpha - \frac{\omega}{2})\omega} \quad \textbf{\textit{Eq. 1}}$$

which is derived from figure 2. The restraint describes the maximum drop (*compression*) of a surface point $S_2[\rho_2,\theta_2,\phi_2]$ relative to a



neighboring surface point $S_1[\rho_1,\theta_1,\phi_1]$ normal to the chord with endpoints $[\rho_1,\theta_2,\phi_2]$ and $[\rho_1,\theta_1,\phi_1]$. It is designed to create a restraint that is uniform at different values for ρ, despite the changing impact of $\Delta\theta$ and $\Delta\phi$ on surface point spacing. The angle α in equation 3 is a user adjustable *compressibility angle* such that:

$$\alpha = \tan^{-1}\left(\frac{compression}{dist}\right)$$

By adjusting α, the MES can be adjusted to gradually change the binding pocket volume (increasing α will decrease binding pocket volume).

The algorithm proposed must adjust the minimum surface such that the compression restraint is satisfied. This is accomplished through creation of an initial surface that satisfies the compression restraint, followed by compression of this surface towards the minimum surface in iterations such that equation

3 is always satisfied. The initial surface is a sphere, centered at *O*, and tessellated at a user-defined resolution (*R*). The radius of the sphere (*rad*) is set such that it encompasses the entire van der Waals surface with allowance for the shell space *S*. The sphere is tessellated at evenly distributed intervals for θ and ϕ such that the resolution is satisfied in an approximate manner:

$$\phi_{\text{int}} = \pi / \left(\text{int} \left(\left(\frac{\pi}{2 \bullet \sin^{-1}\left[R / (2 \bullet rad) \right]} \right) \right) + 1 \right)$$

$$\theta_{\text{int}}(\phi) = 2\pi / \left(\text{int} \left(\left(\frac{2\pi}{2 \bullet \sin^{-1}\left[R / (2 \bullet rad \bullet \sin\phi) \right]} \right) \right) + 1 \right)$$

During tessellation, neighbors to a surface point are stored in the surface point data structure so that the surface restraint can be enforced in a discrete manner. Each surface point gets two top neighbors, two side neighbors, and two bottom neighbors.

After the MES has been calculated, it must be output in some useable form in order to be applied in other programs. Because the MES is intended for application in force-field based algorithms, considering the surface points as *boundary atoms* will allow for a natural incorporation. The repulsive interactions of these boundary atoms can be included in the force field to describe a steric boundary that is similar in nature to that provided by the macromolecule. Therefore, it is beneficial to output the surface points as a set of boundary atoms in Protein Data Bank (PDB) format. For some grid-based applications, it may be desirable to fill all space within the grid outside the MES with boundary atoms such that the only empty space exists within the binding pocket. This will be accomplished by "thickening" the surface up to grid/box boundaries. During thickening, for each surface point S[ρ, θ, ϕ], a new surface point S_2[ρ_2, θ, ϕ] will be added where $\rho_2 = \rho + \gamma$ with γ equal to some specified value in angstroms.

The *boundary atom* representation of the surface poses two problems to the algorithm as described. First, the surface will seem closer to the macromolecule because it is now described by the outer edges of the van der Waals radii of the points, not the points themselves. This can be alleviated easily with a parameter for the radius of a boundary atom. The value for this parameter can then be added to the shell space to correct for the sphere representation.

The second problem is that many algorithms perform optimization where gradient information or energy sampling plays an important role in convergence. The spacing between surface points (boundary atoms) plays a significant role in determining local potentials. Boundary atoms placed too far apart will lead to holes in the surface, while boundary atoms placed too close together will yield unnaturally high potentials that might deceive optimization algorithms. The resolution allows some adjustment of the *initial* spacing, however, this spacing will change during MES compression.

The solution proposed is a clean-up function that removes boundary atoms leading to unnatural energies *after* MES calculation. Two clean-up algorithms are implemented - **Strictclean**(*cleandist*) and **Cleanup**(*cleandist*). **Strictclean** can be used to remove boundary atoms when there is sufficient overlap such that a hole will not be generated. The algorithm for **Strictclean** does not *guarantee* that close contacts will be alleviated, however, it can be used to guarantee that no steric holes exist in the surface. The algorithm removes a boundary atom when a top, bottom, and two side neighbors are within *cleandist* from the atom. The data structure for **Strictclean** involves Boolean priority values such that when a boundary atom is removed, the atom above it in the next "thick" layer will not be removed. **Cleanup**, on the other hand, removes any boundary atoms that are within *cleandist* of the boundary atom in question. Thus cleanup guarantees that no two boundary atoms will be within *cleandist* of each other, but cannot guarantee the absence of steric holes. **Cleanup** also uses Boolean priorities, and both algorithms allow for all boundary atoms

outside of a grid or box to be removed. The choice of clean-up functions is dependent on application.

IMPLEMENTATION – *Algorithm 1*

The algorithm has been implemented in ANSI C++. The default values include a 3Å resolution, 1.5Å shell space, and a 1.52Å cleanup distance. The parameter for adjusting the surface is the compression factor (equal to $\tan(\alpha)$ in equation 1). The default value is 1 ($\alpha=\pi/4$). For most cases, only the PDB structure file and possibly a modified compression factor need to be specified. Atomic radii are assigned based on van der Waals radii from the AMBER force field[14]. The algorithm steps and corresponding time complexities are listed below. For time complexities, *s* represents the number of surface points (whose growth is dependent on macromolecule dimensions) and *p* represents the number of protein atoms. The growth rate listed for calculating the MES assumes that the number of iterations required is small and relatively invariant between macromolecules.

1.	Read in PDB File	O(p)
2.	Calculate *O, rad,* atomic radii	O(p)
3.	Create Initial Surface	O(s)
4.	Calculate Minimum Surface	O(s•p)
5.	Calculate MES	O(s)
6.	*Optional:* Thinkening	O(s)
7.	*Optional:* Cleanup	O(s)
8.	*Optional:* Strict Clean-up	$O(s^2)$
9.	Output Surface PDB File	O(s)

MES GENERATION – *Algorithm 2*

It is desirable that MES calculation can be applied for any shape of binding pocket, leaving no exceptions for which computational programs that utilize the MES will fail. However, testing of the compression algorithm revealed potential problems for a small number of cases in which the binding pocket is shallow and located on a convex portion of the protein exterior. In these cases, a low compression factor is necessary. This is sufficient for creating a boundary for the binding pocket in question, however, it leaves a poorly defined surface around the rest of the protein.

For these cases, an alternative algorithm might suffice, where binding pocket volume is considered to be volume that lies between macromolecule atoms. This is similar in concept to the characterization of a binding pocket in the binding-site identification programs Pocket[15] and LigSite[16]. According to this definition, *any* concave space within the macromolecule will be considered to be binding pocket volume, which will yield vast overestimations in certain cases. Therefore, a binding pocket *diameter* factor is introduced to limit the maximum distance between atoms that are considered to compose binding pocket space.

Considering the potential use for grid-based applications, we start by filling a grid (bounding the entire macromolecule or just the area of interest) with regularly spaced boundary atoms at some *resolution.* This represents the initial surface, which is null. All boundary atoms whose van der Waals radii overlap macromolecule atoms can then be removed. This represents the minimum surface, which is equivalent to the van der Waals surface (at the limit where resolution is infinitely small). From this point, we can create cylindrical segments with endpoints at all atom pairs under the restraint that the length of the segment is less than the *diameter* factor. The radius of the segment is set to the maximum of the radii of the atoms that form the segment endpoints.

Creation of the MES then occurs via the removal of any boundary atoms that overlap with cylindrical segments. The result of the calculations is a "thickened" MES similar to that produced for the compression algorithm above when it is grid parameterized. A single layer MES can be produced easily by removing any boundary atoms that are completely surrounded by other boundary atoms within the grid. Incorporation of shell space from the algorithm

above can be accomplished by adding the shell space parameter to the radii of all cylindrical segments. Cleanup algorithms similar to those implemented in the compression algorithm are also implemented in the grid-based algorithm. The result is an alternative MES algorithm where binding pocket volume is controlled by a diameter factor instead of a compression factor.

IMPLEMENTATION – *Algorithm 2*

The algorithm is implemented as part of the MES generation program listed above. The algorithm steps and corresponding time complexities are listed below. The overall growth rate is determined by MES calculation, which checks each boundary atom with every cylindrical segment, which exists between every atom pair. This poses a serious problem, as for a 5000-atom protein there are 25,000,000 segments. If the grid has 10,000 boundary points, then $2.5 \bullet 10^{11}$ segment intersection evaluations are required. Therefore, steps must be taken to decrease the run time.

First, a sweeping plane[17] is used for minimum surface calculation, such that only those macromolecule atoms lying within a range in the x dimension capable of intersecting a set of boundary atoms lying at a specified x coordinate will be evaluated. Second, in cases where the grid does not encompass the entire macromolecule (i.e. active-site centered grids), all segments that do not pass through the grid are "inactivated". Third, a sweeping plane approach is also used for segment intersection checks such that only those segments passing through the x coordinate of the macromolecule atom center are evaluated. Finally, and most significant to run time, is the inactivation of any segments where the boundary atom overlapping the segment center has been removed in minimum surface calculation. This approximation is justified because it suggests that two smaller segments approximately overlap the inactivated segment.

Neighbor information is again stored during initial surface creation such that cleanup can occur in O(s) time. The clean-up algorithms were implemented in a similar manner to that described for the compression algorithm.

1.	Read PDB File, Sort by x	$O(p \bullet \log p)$
2.	Assign Atomic Radii	$O(p)$
3.	Create Initial Surface	$O(s)$
4.	Calculate Minimum Surface	$O(s \bullet p)$
5.	Calculate MES	$O(s \bullet p^2)$
6.	*Optional:* Remove Thick Layers	$O(s)$
7.	*Optional:* Cleanup	$O(s)$
8.	*Optional:* Strict Clean-up	$O(s^2)$
9.	Output Surface PDB File	$O(s)$

MES EFFICACY

We tested the efficacy of both the compression and segment algorithms for MES generation by creating the surface for 50 different enzyme-ligand complexes for which the binding pockets are known. The PDB codes for the enzymes were 10GS, 1A0J, 1A0L, 1A16, 1A30, 1A42, 1AQL, 1B3N, 1CZI, 1DWB, 1EJN, 1ENU, 1ETR, 1FPP, 1GAI, 1GOS, 1GTX, 1HVR, 1K1I, 1LDG, 1MBI, 1QPN, 1QUR, 1RBP, 1RTF, 1STP, 1TLP, 1ULB, 2CPP, 2ER9, 2HCK, 2IFB, 2MCP, 2XIS, 2YPI, 3PTB, 4DFR, 4FUA, 4HMG, 4PAD, 4STD, 4TMK, 5CNA, 5ER1, 5P2P, 5PAH, 6CPA, 8EST, 8GSS, and 9NSE.

For each case, one subunit was isolated (aside from circumstances when ligand binding involved multiple subunits) and the water was removed. Because hydrogens were not added, united atom van der Waals assignments were used. The ligands were removed and a MES was generated at a range of compression/diameter factors at a shell space of 0 or 2Å. A 1Å resolution was used for the compression algorithm and a 0.5Å resolution was used for the segment algorithm. At each compression/diameter factor, the surface was checked for overlap by placing the ligands back into the binding pocket. Additional statistics were generated based on ligand atom to surface point distances. The surfaces were visually

inspected (using PyMOL rendering software [18]) for each test case at one compression/diameter factor based on ligand atom to surface point distances. Run times were recorded using default parameters for each algorithm on all 50 test cases.

APPLICATION TO FLEXIBLE DOCKING

GAs have been shown to be efficacious for flexible ligand docking in a number of software applications[19-24], typically via optimization of an objective function describing the binding affinity of a ligand. Here, we ask whether or not the efficiency of the GA can be improved by "helping" the algorithm to differentiate ligand conformations that lie outside the binding pocket. Presumably, by forcing high fitness energies onto conformations that lie outside the binding pocket, the populations will quickly converge towards conformations whose atoms are inside the binding pocket. This should limit the fitness evaluation of rotation and conformational degrees of freedom to relevant areas of the macromolecule. This is tested here via modification of AutoDock 3.0 to incorporate a MES boundary during search.

AutoDock 3.0 is a program for flexible ligand docking that can be parameterized to run dockings based on the Metropolis algorithm, genetic algorithm (GA), or Lamarckian genetic algorithm (LGA)[22]. The LGA adds a local search operator parameterized by frequency in order to improve performance at local minima. The program uses a pre-calculated energetic grid (generated by AutoGrid 3.0) for interatomic energy evaluations. The dimensions of the grid determine the range of translations for the search space. For a given docking "job", AutoDock is parameterized to perform a certain number of runs. At completion, the answers from each run are clustered into bins based on similarity measured by RMSD and binding energy.

The MES boundary is incorporated into AutoDock 3.0 by addition of a repulsive Lennard-Jones 12-6 boundary term (ΔG_{MES}) into the free energy equation

$$\Delta G_{MES} = \Delta G_{vdW} \sum \left(\frac{A_{ib}}{r_{ib}^{12}} - \frac{B_{ib}}{r_{ib}^{6}} \right) \quad \text{if } r \le vdW(B) \qquad \textbf{Eq. 2}$$

$$\Delta G_{MES} = 0 \quad \text{if } r > vdW(B)$$

for all ligand atoms i and all surface points b. r represents the distance from the atom center to the surface point. A and B represent atom type parameters as specified in reference 22, ΔG_{vdW} represents an empirically determined coefficient for all van der Waals interactions[22], and $vdW(B)$ represents the surface point van der Waals radius as assigned in the MES algorithm.

This representation was chosen for several reasons. First, it is easy to implement, requiring little change to AutoDock/AutoGrid code. The surface points can be appended as boundary atoms to the macromolecule PDB file and a special atom type can be given to the boundary atom. The force field need only be modified to recognize this atom type, and reject its term in the summation if the energy is negative. Second, it is hoped that it will provide a boundary similar in energetic nature to that given by the macromolecule. We can think of this as an extension of the macromolecule out and around the binding pocket opening.

As an additional parameter, we have fitted AutoDock with an option that allows population seeding such that initial populations can be restricted to translation values that lie within the binding pocket determined by the MES.

Prior to publication, AutoDock was validated via reproduction of structural data from 7 ligand-enzyme complexes previously determined by spectroscopy. We have used these 7 test cases to evaluate the impact of the MES-based search space changes (bound cases) on docking efficiency along with an additional 7 test cases. The testing was performed according to the methods described in reference 22 based on an energetic grid 22.5Å^3 in volume located at the crystallographic ligand structure's center of mass. One potential use of AutoDock involves

docking a ligand when the binding pocket is not known *a priori*. We have tested this application by setting the grid to be the bounding box of the entire enzyme for 2 of the test cases.

The stochastic nature of the GA/LGA makes comparison difficult due to statistical sampling error. We have attempted to reduce this error via the use of random number generator seeds. AutoDock was modified to read in and reproduce random number seeds from other jobs. In this manner, we could compare the effects of the MES when the same initial population was used in the bound case and in the control.

An additional problem with GA analysis, at least in this case, is that the determination of a "correct" answer is somewhat ambiguous. This stems from the fact that the search space is infinite, that structures determined by spectroscopy have significant uncertainty, and that the force-field used here for docking relies on certain simplifications in order to make run-times acceptable. For this analysis, a *reference structure* ligand-binding mode was determined for each test case and assumed to represent the global minimum according to the force-field function. An answer was considered to be correct if the conformation was within 1 Å RMSD from the reference compound and had an energy no more than 1 kcal/mol greater than the reference energy. This left the possibility for binding modes found during analysis with low energy and high RMSD from the reference or low RMSD with the reference and higher energies. Both cases would have significant impact on the results, and therefore analysis for these instances was included in testing.

GA efficiency was determined by analysis of averages of the final number of correct answers, the final best energies, the energies of correct individuals, the final number of conformational cluster bins, the population of the top ten bins, and the CPU time in the bound case required to reach a similar average of correct individuals found in the unbound case.

DOCKING METHODS

AutoGrid 3.0 was modified to recognize the boundary atom type and incorporate it into the force field as described in equation 2. The van der Waals radius and Lennard-Jones parameters for the boundary atom type were set to those for carbon in the AutoGrid parameter files. AutoDock 3.0 was modified to read in seeds from other runs, generate extensive output for post-analysis, and to support population seeding as described above.

The seven cases previously reported (1HVR, 1STP, 2CPP, 2MCP, 3PTB, 4HMG, 4DFR) were used for testing along with seven additional cases (1A16, 1EJN, 1FKF, 2YPI, 4FUA, 4STD, 5CNA). The docking methods, including enzyme and ligand preparation, were identical to those reported in reference 22 aside from a difference in the energy evaluation limit ($2.5 \cdot 10^5$ used here) and number of runs (40 used here). In addition to these 14 test cases, 1HVR and 2CPP were also run using a grid whose dimensions were set according to the bounding box for the enzymes such that the entire proteins were searched during docking (1HVRall and 2CPPall).

For each test case, a reference structure (for evaluation of correct individuals) was generated based on the best conformation found from a docking job consisting of 200 runs and $1.5 \cdot 10^6$

PDB Code	RMSD (Å)	Docked Energy (kcal/mol)	Number Boundary Atoms	Free Bond Torsions
1HVR	0.711	-21.37	438	10
1HVRall	0.711	-21.37	6698	10
1STP	0.599	-9.65	689	5
2CPP	0.844	-7.42	7	0
2CPPall	0.844	-7.42	9884	0
2MCP	0.956	-5.94	400	4
3PTB	0.539	-8.48	393	0
4DFR	1.051	-11.76	396	11
4HMG	0.767	-7.58	765	10
1A16	1.286	-8.61	437	4
1EJN	0.570	-11.78	382	8
1FKF	1.415	-14.13	556	15
2YPI	0.995	-6.05	476	3
4FUA	0.764	-6.04	668	5
4STD	0.988	-6.62	313	4
5CNA	0.969	-7.41	768	6
8GSS	1.075	-10.21	591	11

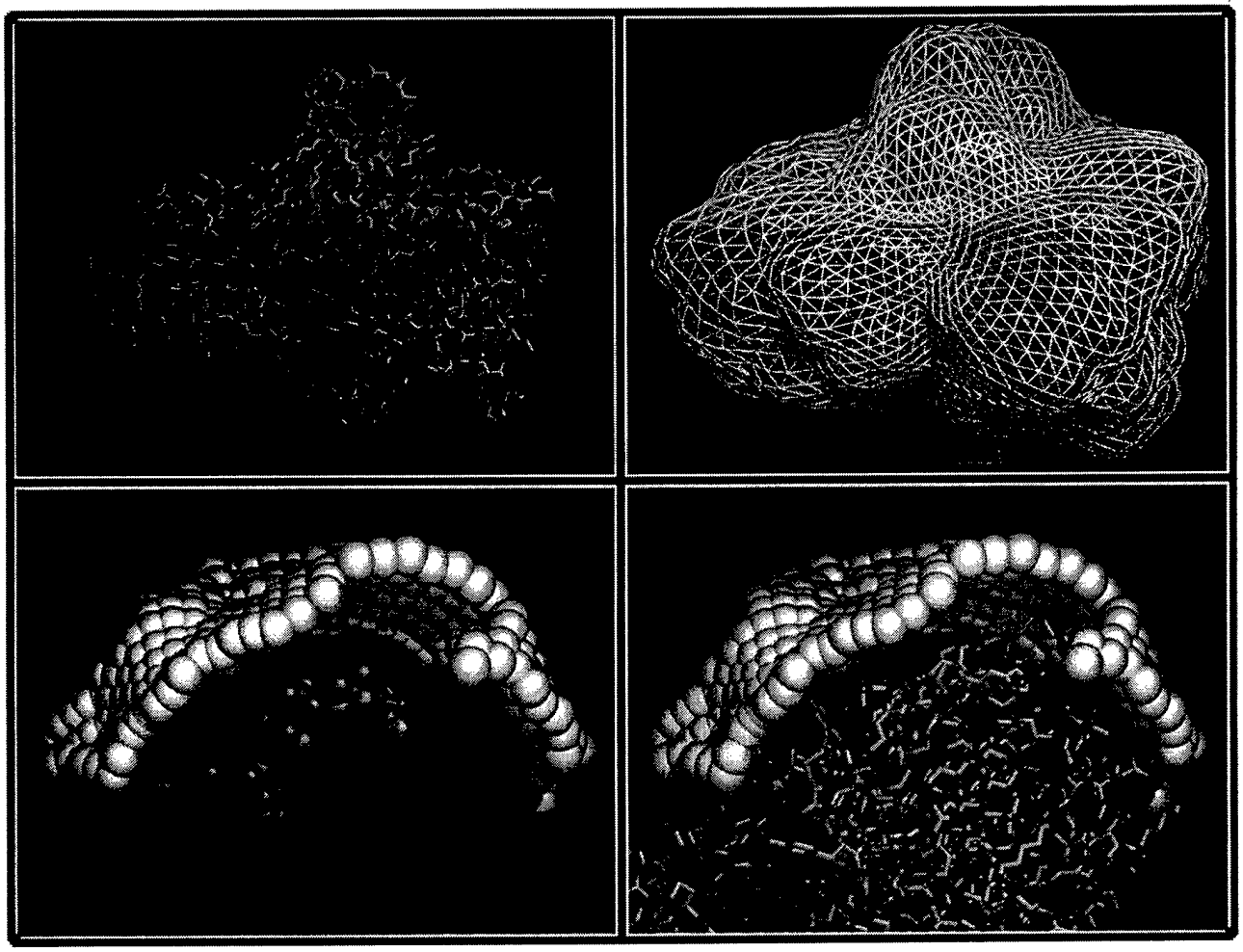
energy evaluations. The RMSDs of each reference structure from the crystallographic conformations, along with the docking energy and number of rotatable bonds considered by AutoDock are listed in table I. The MES for each test case was generated, *with the ligand in place,* using the compression algorithm with default parameters (compression factor of 1). The surface was thickened and cleaned according to the dimensions of the grids used in AutoGrid 3.0. The number of boundary atoms influencing each grid is listed in table I.

For each enzyme, we created a "bound case" by appending boundary atoms influencing the grid onto the protein input file. The bound and control cases were run in an identical manner, using the same random number seeds. Four job types were considered including the control, the control with population seeding, the bound case, and the bound case with population seeding. For each job type consisting of 40 runs, 40 jobs were executed to get a measure of reproducibility. After completion, the output files were parsed with a program implemented in C++ that performed RMSD analysis and generated statistics for the jobs at each generation.

Results

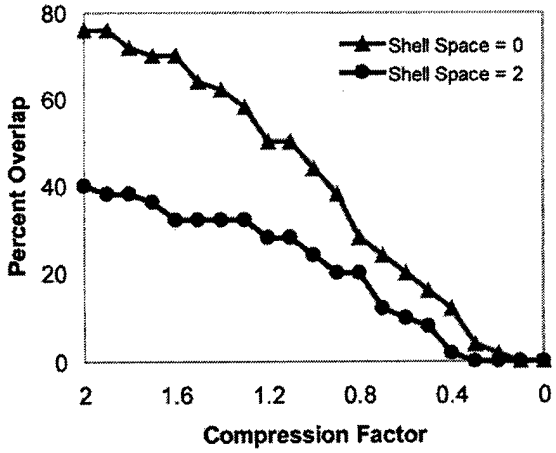
MES EFFICACY

The efficacies of the proposed algorithms were evaluated on a set of 50 enzyme-ligand complexes for which the binding sites are known. There should never be a problem with

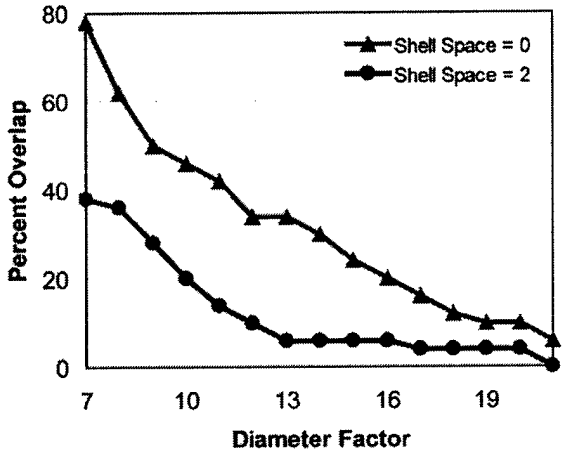


MES generation when the structure of a ligand that fully occupies the binding pocket is known. In this case, the minimum surface of the macromolecule with the ligand bound could represent the MES. However, it is desirable that the surface can be generated for macromolecules in which the ligand binding modes are unknown, or do not fully occupy the binding pocket. We therefore generated test cases by removing the ligand from each structure, generating the MES at different values for each algorithms adjustable factor, and replacing the ligand to test for MES overlap (figure 3). Because it is difficult to quantitate the "success" of a MES calculation due to the vastly diverse topologies of proteins and the potential for specific experimental needs, visual inspection played an important role in algorithm testing.

The results from the compression algorithm are shown in figure 4. The algorithm worked well for most of the test cases. For the majority of cases, a compression factor of 1 led to calculation of an appropriate surface and binding pocket volume. However, for 1GAI, 1K1I, and 10GS, low compression factors were required to prevent overlap with ligand binding modes. The proteins from these crystallographic structures contain binding pockets that are shallow and near the protein exterior. While the surfaces generated create adequate boundaries for the binding pockets in question, the surfaces are poorly defined around the rest of the exterior of the proteins. For certain foreseeable applications, including binding pocket identification or ligand screening involving the entire protein, this might generate problems.



This prompted the development of the alternative algorithm for MES generation described above. This algorithm was tested using adjustment of the diameter factor as shown in figure 5. The algorithm worked well for the problematic test cases described above. In order to compare the two algorithms, distances between ligand atoms *exposed* to the MES surface and MES surface points were calculated for each test case using the two algorithms with varied adjustable factors. The distances were compared at the adjustable factor for each algorithm that resulted in the same percent overlap reported in figures 4 and 5. The averages, maximums, minimums, and standard deviations for the distances were strikingly similar (data not shown) between the two algorithms, aside from small differences at low compression factors.



Based on this data, there is little reason to choose one algorithm over the other for most test cases. The grid-based algorithm suffers to a larger degree from discrete error and for several of the test cases it is difficult to make small adjustments to binding pocket volumes. We have therefore chosen to make the compression algorithm the default, although the software for MES generation allows for use of the other algorithm for special circumstances.

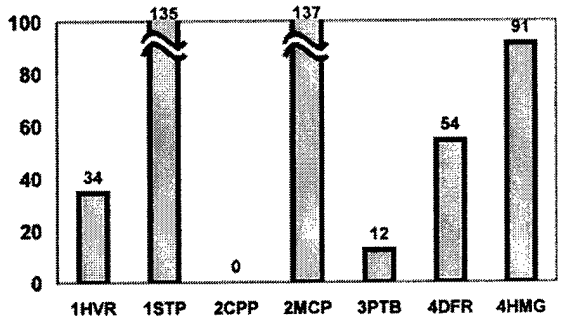
RUN TIMES

The run times were recorded on a 1.9 GHz Pentium 4 Dell Precision WorkStation 340 running Red Hat Linux 7.0. Run times were recorded for all of the test cases using the default parameters and the strict-cleanup processing. The average run time for the compression algorithm was 1.5 seconds. MES generation for all test cases required under 4 seconds, excluding 4HMG which took 11.8 seconds due to the large number of surface points required for the initial surface to encompass the "baseball bat-shaped" enzyme. The MES calculation step of the algorithm only requires thousandths of a second, making visual adjustment of the compression factor plausible in graphic rendering programs. At a resolution of 1.5 angstroms, the grid-based algorithm run-times were comparable. The average run time was 1.6 seconds and the maximum time required was 3.88 seconds.

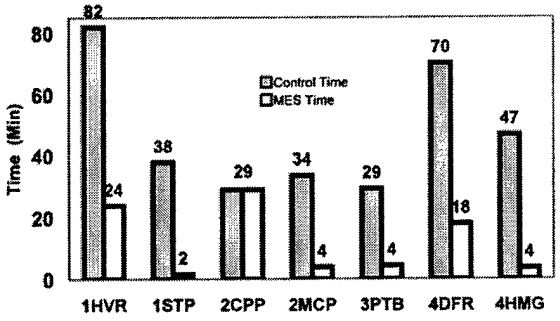
APPLICATION TO DOCKING

We have suggested that flexible docking efficiency can be improved for the GA and LGA by enforcing MES boundaries such that the only empty volume lies within the binding pocket, and nowhere else outside the macromolecule. This was tested by comparing cases with and without the boundary in place, using the same initial population and random generator seeds. For the GA, the 7 test cases reported for AutoDock validation[22] were used for testing.

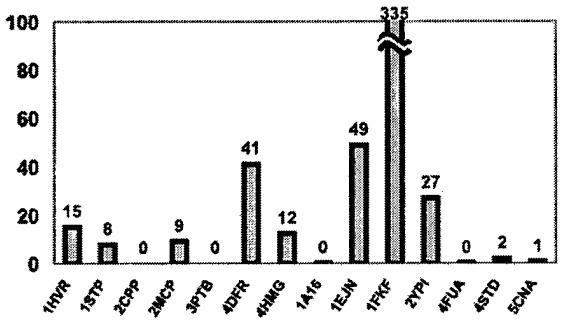
For most cases, substantially improved results were obtained for the GA when MES boundaries were enforced. As expected, results from the 2CPP test case were invariable due to the fact that the MES surface did not pass through the energetic grid (only 7 boundary atoms influenced grid energies – table I). On average, there were over twice as many correct individuals for two of the test cases and significant increases in the others (figure 6). In the extremes of improvement, there was, based on averages of 40 jobs (1600 runs), a 137% increase in correct individuals (2MCP), a 0.65 kcal/mol decrease in best energy (4DFR), a 0.05 kcal/mol decrease in energy of correct individuals (1STP), a 31% decrease in number of conformational bins (2MCP), and a 42% increase in population of the top ten bins (2MCP) [each value taken as the best from the 7 test cases].

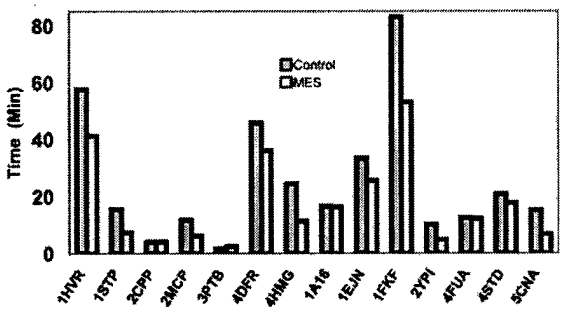


We also looked at the percent decrease in run time that could be obtained using MES boundaries in order to reach results similar to those produced by the control. For example, the maximum number of correct individuals for the 1STP case was reached at generation 5000 (average time of 37.8 minutes) with an average of 3.01 correct. In the bound case, an average of 3.13 correct individuals was reached at generation 200 (average time of 1.65 minutes). This gives a 96% reduction in the run time required to reach the same results as the control. The run times required to reach similar results are illustrated in figure 7.



For the LGA, an additional 7 enzyme-ligand complexes were used for testing. The results show significant, however, less dramatic improvements compared to those seen in the GA. In the extremes of improvement there was, on average, a 335% increase in correct individuals (1FKF), a 2.21 kcal/mol decrease in best energies (1FKF), a 0.16 kcal/mol decrease in energy of correct individuals (1STP), a 63% decrease in conformational bins (1STP), a 79% increase in population of the top ten bins (1FKF), and a 55% decrease in the run time required to reach similar results (5CNA). The percent increase in correct individuals and run time comparisons for similar results are illustrated in figures 8 and 9 respectively.





The 3PTB test case, which represents the easiest optimization for flexible docking out of the test cases here, resulted in a slightly longer time required to reach the control results (2.28 minutes instead of 1.63 minutes). 2CPP, of course, resulted in no noticeable differences between bound and control cases. For all other test cases in the GA and LGA, improvements in all aspects of the results as examined here were seen. For the LGA, there is some correlation between the degree of improvement and the number of boundary atoms influencing the grid along with the degree of flexibility in the ligand. Indeed, the most substantial improvement was seen for the 1FKF test case (with 15 degrees of flexibility and 556 boundary atoms influencing the grid). This trend is likely complicated by other factors such as binding pocket shape and energetics.

When the binding pocket for a specific ligand is not known *a. priori*, AutoDock might be used to search an entire macromolecule for potential ligand binding sites. We tested the influence of the MES for such instances by searching the entire enzyme via LGA for both the 1HVR and 2CPP test cases. The results are summarized in table II. The control 1HVR case failed with only 0.3 correct individuals on average. A substantial improvement was seen when the MES was enforced – an average of 13.7 correct individuals. The 2CPP test case involved a rigid ligand with no rotatable bonds. Nonetheless, a 48% increase in correct individuals was observed in the bound case. Therefore, a vast improvement is expected for cases in which the entire macromolecule is searched for ligand binding.

PDB Code	% Increase in Correct Individuals	% Decrease in Run Time	Difference in Best Energy (kcal/mol)	Difference in Correct Energy (kcal/mol)
1HVRall	4864	75	-2.66	-0.19
2CPPall	48	71	-0.36	0.00

The tests were performed on a 8 processor SGI Origin using 300MHz MIPS R12000 CPUs and running Irix 6.5. It is also noteworthy that no optimization of genetic algorithm parameters for use with MES boundaries was performed. It is expected that adjustment of parameters for MES incorporation will offer further improvement of LGA results, however these experiments have not been performed (see below).

The idea of population seeding such that the center atom of every individual in a population lies within the binding pocket was also tested. The results were indistinguishable from the

controls in most cases, presumably due to quick convergence of translation variables towards the binding pocket in control cases. Therefore, the data pertaining to population seeding was left out.

Discussion

Numerous algorithms have been described for the characterization and identification of binding pockets. Most of these algorithms define the binding pocket by evaluating a descriptor on either a set of grid points[10, 15, 16, 25, 26] or on a set of probes[27-29] (i.e. spheres placed tangential to protein atoms or molecular fragments placed around the macromolecule). The descriptors used by these algorithms have included probe interaction energies[29], accessibility parameters (surface accessibilities[25], burial counts of nearby protein atoms[27], or volumes that become inaccessible due to atom fattening[10]), an angular condition that identifies spheres within concave regions[28], or the presence of surrounding protein atoms collinear with the grid point[15, 16, 26]. Perhaps the most elegant solution was offered by Liang, Edelsbrunner, and colleagues[30]. Using computational geometry tools, they characterize the macromolecule as a weighted Delaunay triangulation. The "empty" Delaunay tetrahedra can then be analyzed and possibly merged to characterize the binding pocket.

These algorithms were not chosen for MES generation either because they fail to identify the sea level of the binding pocket, they are inapplicable to a significant proportion of macromolecules, or the calculated sea level is difficult or impossible to adjust. The binding pocket is an ambiguous concept. In theory, we can conceive of a molecule that can bind to and fill any concave region of a macromolecule. Within this, there are practical limits governed by necessary physiologic properties of molecules and synthetic limits. These limits might be used to characterize what "usually" composes a binding pocket; however, it is essential that these methods be adjustable according to experimental need or algorithm failures.

For example, the method employed by Liang et al. (discrete flow method), defines the sea level of the binding pocket as the region where paths into the pocket become narrower than the binding pocket itself. This definition is advantageous in that it allows automatic characterization of binding pockets without user parameters and provides criteria for statistical comparison of enzymes on a large scale. However, it precludes delineation of binding sites that do not contain narrow binding site openings (a minor, yet significant proportion of enzyme binding sites). Additionally, although it is not reported in their experiments, the discrete flow method would seem to offer the potential for significant underestimation of binding site volume.

We have presented two novel algorithms for calculation of binding pocket sea levels, and consider them to be advantageous in that binding pocket volume can be easily and gradually adjusted. The MES can be visualized as a net or spacefill model (as shown in figure 3) in any molecular visualization program, affording visual adjustment of binding pocket sea level. Additionally, the MES approach allows for efficient boundary enforcement in force-field based algorithms. One drawback of the algorithms is their inability to automatically characterize what "usually" composes binding pocket space. While a compression factor of 1 works for the majority of cases, a significant fraction requires adjustment. Therefore, future work will consider the use of supervised learning techniques to aid the user in parameterization.

The "boundary atom" representation of the MES allows for the identification and characterization of binding pockets. Each range of continuous empty space within the MES represents a binding pocket. Using the flood-fill algorithm[31] with a probe sphere of some radius, the volume, surface area, and solvent accessible

atoms of each binding pocket can be easily calculated. These descriptors may be used for identification of potential binding pockets when a ligand-binding mode is not known *a priori*. Additionally, binding pocket descriptors might be useful in the automatic parameterization of stochastic docking and *de novo* design algorithms. For example, the population size, number of generations, and mutation rate necessary to yield a correct answer might be predicted based on the ligand degrees of freedom, binding pocket volume, surface area, and number of hydrogen bonds.

Perhaps the most useful application of MES calculation is for characterization of the search space for ligand docking and *de novo* design algorithms. The use of active-site centered spheres, boxes, or atoms interacting with a crystallographic ligand for generating the search space might be appropriate for validation – we already know where the ligand is supposed to go. However, it is inappropriate for many useful applications, either because the methods include search space that is not relevant, or miss search space that might be useful for ligand binding.

One example can be seen from the crystal structure of urokinase plasminogen activator (1EJN)[32], a serine protease utilizing binding pocket interactions with 6 amino acid residues surrounding the scissile bond. The protein is crystallized with an inhibitor occupying only a portion of the relatively large binding pocket. The default dimensions of the sphere or box used by most programs will not incorporate the entire binding pocket. Using the MES and the flood-fill algorithm as described above, however, all of the interactions sites with potential for ligand binding can be identified, and the minimum dimensions of a box or sphere needed to encompass the entire binding pocket can be calculated easily. Application of the MES for these cases (including binding site identification and characterization) has been implemented in a program called Binding Pocket Surveyor. The details will be described in a separate paper

Lead optimization algorithms have been described that perform systematic addition of molecule fragments or peptide residues to a seed molecule[7, 8]. An important bound to the combinatorial explosion inherent in these types of algorithms is the binding pocket boundary, defined by the repulsive interactions of the macromolecule atoms with those of the growing ligand[8]. Therefore, MES calculation might be especially useful as a means to limit ligand growth in a computationally efficient manner.

We have demonstrated the use of the MES to improve the efficiency of flexible ligand docking in the genetic algorithms applied in AutoDock. The stochastic nature of the GA and the expensive CPU cost for fitness evaluations makes GA analysis difficult. Exact mathematical models of the GA are limited to simple, impractical applications[33], and therefore it is difficult to answer such questions as "How can GA efficiency be improved?" In the spirit of traditional GA theory, however, it might be expected that the population would quickly converge towards translations within or near the binding pocket, as these individuals have binding affinities improved by many orders of magnitude. Indeed, observation of best energies as the GA progresses reveals a rapid initial drop in energies as steric overlap is alleviated. Based on these considerations, MES boundary enforcement would seem to have little impact on docking efficiency.

The key observation, however, is that ligand rotation and conformational degrees of freedom are not independent from translation variables. The optimum ligand conformation at one translation is almost certainly different from that at another translation. However, the fitness value makes no attempt to distinguish between translation and conformational fitnesses. It therefore seems reasonable that during convergence of translation variables towards the binding pocket, there is an associated convergence towards values for the rotation and conformational degrees of freedom that are not

relevant to conformations where ligand atoms lie entirely within the binding pocket.

It is therefore presumed that enforcing MES boundaries during ligand docking helps to prevent the loss of relevant conformational variables during early convergence by distinguishing in an energetic fashion between those conformations that exist inside or outside the binding pocket. The LGA incorporates local search, which may help to reintroduce relevant conformational values that were lost during translation convergence. This would explain the less dramatic improvement seen for MES incorporation during LGA search. However, this also suggests that the impact of local search is lessened in the bound cases, and that docking efficiency in these cases can be improved by lowering the frequency of expensive local searches. This seems to be the circumstance for the test cases 1HVR and 1STP (in data not reported), however, no full optimization involving all test cases was performed.

Incorporation of MES boundaries for flexible docking had the most significant impact when the entire enzyme was searched for ligand binding modes. This may be useful for elucidating pathways that might result from a compound's inhibition of one potential enzyme, or allosteric activation of another. In cases with flexible ligands, it is expected that extensive run-times would be required in order to obtain a correct answer when MES boundaries are not used.

Acknowledgements

The authors would like to thank the developers of AutoDock and PyMOL for making their source code freely available via the Internet.

References

1. Ewing, T. J.; Makino, S.; Skillman, A. G.; Kuntz, I. D. J Comput Aided Mol Des 2001, 15, 411-28.
2. Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. J Mol Biol 1996, 261, 470-89.
3. Bohm, H. J. Perspect Drug Discov 1995, 3, 21-33.
4. Clark, D. E.; Frenkel, D.; Levy, S. A.; Li, J.; Murray, C. W.; Robson, B.; Waszkowycz, B.; Westhead, D. R. J Comput Aided Mol Des 1995, 9, 13-32.
5. Trosset, J.; Scheraga, H. J Comput Chem 1999, 20, 412-427.
6. Pang, Y. P.; Perola, E.; Xu, K.; Prendergast, F. G. J Comput Chem 2001, 22, 1750-1771.
7. Luo, Z.; Wang, R.; Lai, L. J Chem Inf Comp Sci 1996, 36, 1187-1194.
8. Budin, N.; Majeux, N.; Tenette-Souaille, C.; Caflisch, A. J Comput Chem 2001, 22, 1956-1970.
9. Laskowski, R. A. J Mol Graphics 1995, 13, 323-&.
10. Kleywegt, G.; Jones, T. Act Cryst D 1994, 50, 178-185.
11. Lee, B.; Richards, F. M. J Mol Biol 1971, 55, 379-400.
12. Coxeter, H. Introduction to Geometry; Wiley: New York, 1969.
13. Richards, F. M. J Mol Biol 1974, 82, 1-14.
14. Weiner, S.; Kollman, P.; Nguyen, D.; Case, D. J Comput Chem 1986, 7, 230-252.
15. Levitt, D.; Banaszak, L. J Mol Graphics 1992, 10, 229-234.
16. Hendlich, M.; Rippmann, F.; Barnickel, G. J Mol Graphics 1997, 15, 359-&.
17. De Berg, M.; Van Kreveld, M.; Overmars, M.; Otfried, S. Computational Geometry; Springer: 2000.
18. Delano, W. L. The PyMOL Molecular Graphics System; DeLano Scientific: San Carlos, 2002.
19. Hou, T.; Wang, J.; Chen, L.; Xu, X. Prot Eng 1999, 12, 639-647.
20. Oshiro, C.; Kuntz, I.; Dixon, J. J Comput-Aided Mol Des 1995, 9, 113-130.
21. Jones, G.; Willett, P.; Glen, R.; Leach, A.; Taylor, R. J Mol Biol 1997, 267, 727-748.
22. Morris, G. M., Goodsell, D.S., Halliday, R.S., Huey, R., Hart, W. E., Belew, R. K., and Olson, A.J. J Comput Chem 1998, 19, 1639-1662.
23. Taylor, J.; Burnett, R. Proteins 2000, 41, 173-191.
24. Yang, J.; Kao, C. J Comput Chem 2000, 21, 988-998.
25. Stahl, M.; Bur, D.; Schneider, G. J Comput Chem 1999, 20, 336-347.
26. Exner, T.; Keil, M.; Moeckel, G.; Brickmann, J. J Mol Model 1998, 4, 340-343.
27. Brady, G. P., Jr.; Stouten, P. F. J Comput Aided Mol Des 2000, 14, 383-401.
28. Kuntz, I.; Blaney, J.; Oatley, S.; Langridge, R.; Ferrin, T. J Mol Biol 1982, 161, 269-288.
29. Ruppert, J.; Welch, W.; Jain, A. N. Prot Sci 1997, 6, 524-533.
30. Liang, J.; Edelsbrunner, H.; Woodward, C. Prot Sci 1998, 7, 1884-97.
31. Foley, J. D.; Van Dam, A. Fundamentals of Interactive Computer Graphics.; Addison-Wesley: Reading, 1983.
32. Sperl, S.; Jacob, U.; Arroyo De Prada, N.; Sturzebecher, J.; Wilhelm, O. G.; Bode, W.; Magdolen, V.; Huber, R.; Moroder, L. Proc Natl Acad Sci U S A 2000, 97, 5113-8.
33. Nix, A. E.; Vose, M. D. Ann Math Artif Intel 1991, 5, 79-88.

Captions:

FIGURE 1. Vector diagram illustrating the calculation of the minimum surface. The value for ρ based on macromolecule atom M_j for the surface point in question is given by ρ_j.

FIGURE 2. Illustration of the surface restraint $[\Delta\rho(S_1,S_2,\omega)]$ (equation 1) for the MES based on two neighboring surface points. The compression factor used by the software is expressed in terms of *distance* divided by *compression* (tan α).

TABLE I. Reference ligand binding mode data

FIGURE 3. MES generation for lactate dehydrogenase (1LDG). *Upper left:* Ternary complex of enzyme with substrate and cofactor. *Upper right:* MES surface (white mesh) generated using compression factor of 1 for the apoenzyme (without substrates). *Lower left:* Zoomed view of crystallographic substrate and cofactor with nearby surface points represented using a spacefill model (white). *Lower right:* Zoomed view with the enzyme present.

FIGURE 4. Percentage of enzyme-ligand complexes exhibiting steric overlap with the MES at varying compression restraints (*n=50).*

FIGURE 5. Percentage of enzyme-ligand complexes exhibiting steric overlap with the MES at varying diameter restraints (*n=50).*

FIGURE 6. Percent increase in the average number of successful runs in the GA when the MES boundary is enforced (*n=40*).

FIGURE 7. Average run times required to reach a similar average of successful runs in the GA for the control and MES cases (*n=40*).

FIGURE 8. Percent increase in the average number of successful runs in the LGA when the MES boundary is enforced (*n=40*).

FIGURE 9. Run times required to reach a similar average of successful runs in the LGA for the control and MES cases (*n=40*).

TABLE II. Comparison between control and MES data for search of the entire enzyme.